Office of Technology Assessment

Congressional Board of the 98th Congress

MORRIS K. UDALL, *Arizona, Chairman*

TED STEVENS, *Alaska, Vice Chairman*

Senate

ORRIN G. HATCH
Utah

CHARLES McC. MATHIAS, JR.
Maryland

EDWARD M. KENNEDY
Massachusetts

ERNEST F. HOLLINGS
South Carolina

CLAIBORNE PELL
Rhode Island

House

GEORGE E. BROWN, JR.
California

JOHN D. DINGELL
Michigan

LARRY WINN, JR.
Kansas

CLARENCE E. MILLER
Ohio

COOPER EVANS
Iowa

JOHN H. GIBBONS
(Nonvoting)

Advisory Council

CHARLES N. KIMBALL, *Chairman*
Midwest Research Institute

EARL BEISTLINE
University of Alaska

CHARLES A. BOWSHER
General Accounting Office

CLAIRE T. DEDRICK
California Land Commission

JAMES C. FLETCHER
University of Pittsburgh

S. DAVID FREEMAN
Tennessee Valley Authority

GILBERT GUDE
Congressional Research Service

CARL N. HODGES
University of Arizona

RACHEL McCULLOCH
University of Wisconsin

WILLIAM J. PERRY
Hambrecht & Quist

DAVID S. POTTER
General Motors Corp.

LEWIS THOMAS
Memorial Sloan-Kettering Cancer Center

Director

JOHN H. GIBBONS

This is an OTA Technical Memorandum that has neither been reviewed nor approved by the Technology Assessment Board

SCIENTIFIC VALIDITY OF POLYGRAPH TESTING

A RESEARCH REVIEW AND EVALUATION

A TECHNICAL MEMORANDUM

NOVEMBER 1983

Technical Memoranda are issued by OTA on specific subjects analyzed in recent OTA reports or on projects presently in process at OTA. They are issued at the request of Members of Congress who are engaged in committee legislative actions which are expected to be resolved before OTA completes its assessment.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D. C. 20510

J952-68

Recommended Citation:

Scientific Validity of Polygraph Testing: A Research Review and Evaluation—A Technical Memorandum (Washington, D.C.: U.S. Congress, Office of Technology Assessment, OTA-TM-H-15, November 1983).

Library of Congress Catalog Card Number 83-600618

For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402

Foreword

This technical memorandum presents the results of the Office of Technology Assessment's (OTA) review and assessment of the scientific evidence on the validity of polygraph testing. Conducted at the request of Rep. Jack Brooks, Chairman, House Committee on Government Operations, and Rep. Frank Horton, the Ranking Minority Member, the OTA memorandum is intended to assist the committee in its deliberations on proposed changes in polygraph use by the Federal Government.

As requested, OTA has limited this technical memorandum to issues directly related to the scientific validity of the polygraph. OTA did not consider utility, privacy, constitutional, and ethical issues, among others that have been raised in the debate over polygraph testing.

We first discuss the various types of polygraph testing procedures and ways in which the polygraph is used, and then summarize the judicial, legislative, and scientific controversy over polygraph testing validity. Next, we review and evaluate both prior reviews of the scientific research on polygraph validity and the individual research studies. Finally, we discuss the range of factors that may affect polygraph validity and the possibilities for future research, and present OTA's conclusions about the scientific evidence for current and proposed Federal Government polygraph use.

In preparing this memorandum, OTA has drawn on research information available from a wide variety of sources, including the major Federal Government polygraph users, the American Polygraph Association, various private polygraph practitioners, and polygraph researchers both in the United States, and abroad.

In addition to the members of the project advisory panel, this memorandum benefited from the consultation and review of a large number of persons in the Federal Government, universities, and the polygraph community. It is, however, solely the responsibility of OTA, not those who advised and assisted us in its preparation.

Polygraph Validity Advisory Panel

Edward S. Katkin, *Chairman*
Professor of Psychology, State University of New York at Buffalo

Joseph P. Buckley
President
John E. Reid & Associates

Robert Edelberg
Professor of Psychiatry and Psychology
UMDNJ-Rutgers Medical School

Frank Horvath
Associate Professor
School of Criminal Justice
Michigan State University

David T. Lykken
Profesor of Psychiatry and Psychology
University of Minnesota Medical School

Gail J. Povar
Assistant Professor of Medicine and Health Care Sciences
The George Washington University Medical Center

Steve Pruitt
Director of Congressional Affairs
Public Employees Department, AFL-CIO

Christopher H. Pyle
Associate Professor of Politics
Mt. Holyoke College

David C. Raskin
Professor of Psychology
University of Utah

Harold Sigall
Professor of Psychology
University of Maryland

George B. Trubow
Professor of Information Law and Policy
The John Marshall Law School

Althea M. I. Wagman
Research Associate of Psychiatry
Maryland Psychiatric Research Center
University of Maryland School of Medicine

OTA Polygraph Validity Project Staff

John H. Gibbons, *Acting Assistant Director, OTA*
*Health and Life Sciences Division**

Clyde Behney, *Health Program Manager*

Fred B. Wood, *Project Director*
(Communication and Information Technologies Program (CIT))

Leonard Saxe, *Principal Investigator and Author (Boston University)***

Denise Dougherty, *Co-author and Analyst (Health)*†

Theodore Cross, *Co-author (Boston University)*††

Jack Langenbrunner, *Analyst (Health)*

Katherine Locke, *Research Assistant (Health)*

Administrative Support

Ginny Cwalina, *Administrative Assistant (Health)*
Elizabeth Emanuel, *Administrative Assistant (CIT)*
Jennifer Nelson, *Secretary (Health)*
Shirley Gayheart, *Secretary (CIT)*

Other Contributors

Michael Saks, *Boston College*
Daniel Ozer, *Boston University*
Yoram Bar-Tal, *Boston University*
Mary Beasley, *Boston University*
Marie Calabrese, *Boston University*
Molly Zane, *Boston University*

OTA Publishing Staff

John C. Holmes, *Publishing Officer*

John Bergling   Kathie S. Boss   Debra M. Datcher   Joe Henson
Glenda Lawing   Linda A. Leahy   Cheryl J. Manning

---

*Since September 1983; H. David Banta served as Assistant Director until August 1983.
**Assistant Professor of Psychology, Boston University.
†OTA Analyst since September 1983; Boston University Graduate Research Assistant until August 1983.
††Graduate Research Assistant, Boston University.

Acknowledgments

OTA acknowledges the contribution of the following agencies and individuals that provided information, advice, and/or substantive reviews of draft materials:

Federal Government Agencies

Department of Defense (Office of Deputy Under Secretary of Defense for Policy, Army, Air Force, Navy, Marines, National Security Agency)
Department of Justice (Federal Bureau of Investigation, Drug Enforcement Administration)
Department of State
Department of the Treasury (Secret Service, Bureau of Alcohol, Tobacco, and Firearms)
Postal Service
Office of Personnel Management
Central Intelligence Agency
Director of Central Intelligence (Security Committee)

Individuals

Maynard Anderson, Department of Defense
Norman Ansley, National Security Agency
Gordon H. Barland, Barland & Associates
Antonio S. Barrio, Central Texas College System
Louise Becker, Congressional Research Service
William H. Bell, Department of Defense
B. F. Bloomingburg, Naval Investigative Service
Irv Boker, General Accounting Office
Robert A. Brisentine, Army Criminal Investigation Command
Henry L. Canty, Consultant
Richard P. Clayberg, SRI International
Eileen Correa, Veterans Administration Medical Center
Ron Decker, Army Military Policy School
Jack Donnelly, Department of Defense
William R. Fedor, Department of Defense
Michelle Fine, University of Pennsylvania
Robert J. Gatchel, University of Texas at Dallas
Fitz Godwin, Army
Lewis R. Goldberg, University of Oregon
Gisli Gudjonson, London Institute of Psychiatry
David L. Hammond, Psychologist
James E. Hardy, Air Force Office of Special Investigations
Fred Hegge, Walter Reed Army Institute of Research
Eric J. Holden, Baker, Holden & Associates
Charles Robert Honts, University of Utah
William G. Iacono, University of British Columbia
Michel Pierre Janisse, University of Manitoba
Richard C. Johnson, Polygraph, Inc.
Scott Kingsley, National Security Agency
Benjamin Kleinmuntz, University of Illinois at Chicago Circle
Brian E. Lynch, Canadian Police College
Lynn P. Marcy, American Polygraph Association
Paul K. Minor, Federal Bureau of Investigation
L. D. Noland, Army Intelligence and Security Command
Jesse Orlansky, Institute for Defense Analysis
John A. Podlesny, Federal Bureau of Investigation
Roland Radloff, National Science Foundation
Richard K. Riegelman, George Washington University Medical Center
John Roberts, Civil Liberties Union of Massachusetts
Robert Rosenthal, Harvard University
Akihiro Suzuki, Japanese Institute of Police Science
Julian J. Szucko, University of Illinois at Chicago Circle
William A. Thomas, American Bar Foundation
Howard W. Timm, Southern Illinois University of Carbondale
H. Herbold Wooten, Institute of Applied Polygraph Science
Lawrence S. Wrightsman, University of Kansas

Contents

Chapter *Page*

1. SUMMARY
   - Introduction 3
   - Federal Polygraph Use 3
   - Federal Polygraph Policy Changes 3
   - Polygraph Validity 4
   - Findings 4
     - Personnel Security Screening 4
     - Criminal Investigations 5
     - False Negatives/Countermeasures 5
     - False Positives 5
     - Voluntary v. Involuntary 6
     - Polygraph Theory 6
     - Further Research 6
   - Chapter-by-Chapter Overview 6
   - Conclusions 7

2. VARIETIES OF POLYGRAPH TESTING AND USES 11
   - Introduction 11
   - Polygraph Instrument 11
   - Types of Testing Procedures 12
     - The Pretest Interview 12
     - Types of Questions 14
     - Polygraph Question Techniques 17
   - Post-Test Interview 23
   - Uses of Polygraph Testing 23
     - Current Use 23
   - Conclusions 25

3. CONTROVERSY OVER POLYGRAPH TESTING VALIDITY 29
   - Introduction 29
   - Judicial Reviews 29
     - Polygraph Findings as Evidence 30
     - Laws Regulating Polygraphs in Employment Settings 31
   - Federal Debate Over Polygraph Validity 32
     - The 1960's 32
     - The 1970's 34
     - The 1980's 35
     - Draft Revisions to DOD 5210.48 36
     - NSDD-84 36
     - Administration Polygraph Policy 36
   - Scientific Validity and Polygraph
     - Research Reviews 37
     - Definitions of Scientific Validity 37
     - Research Design 38
     - False Positives and Negatives 38
     - Reviews of Polygraph Validity 39
   - Conclusions 43

4. REVIEW AND ANALYSIS OF POLYGRAPH FIELD STUDIES 47
   - Introduction 47
   - Study Selection 47
     - Studies Excluded 48
     - Studies Included 49
   - Coding 51
   - Findings and Discussion 51
     - Variation Among Studies 52
     - Studies Using Panel Criterion and Examiners' Decisions 52
     - Studies Using Confession as a Criterion and Blind Evaluations 54
     - Studies Using Judicial Outcomes and Original Examiners' Results 56
   - Other Considerations 56
   - Conclusions 58

5. REVIEW AND ANALYSIS OF POLYGRAPH ANALOG STUDIES 61
   - Introduction 61
   - Characteristics of Analog Studies 61
     - Study Selection 62
     - Description of Studies 65
   - Control Question Technique 65
     - University of Utah Studies 65
     - Other Studies 68
   - Concealed Information Tests 72
     - Lykken 72
     - Davidson 73
     - Podlesny and Raskin 73
     - Giesen and Rollison 73
     - Balloun and Holmes 74
     - Bradley and Janisse 75
     - Timm 75
   - Preemployment Screening 75
     - Correa and Adams 75
     - Barland 76
   - Findings 78

6. FACTORS AFFECTING POLYGRAPH EXAMINATION VALIDITY 83
   - Introduction 83
   - Polygraph Examiner, Subject, and Setting 83
     - Examiner 83
     - Subjects 84
     - Setting 86
   - Countermeasures 87
     - Physical 87

Drugs 88
Hypnosis/Biofeedback 89
Mental 89
Research Implications of Factors Affecting Validity 91
Theory 91
Scoring 91
Question Techniques 91
Countermeasures 91
Comparison With Other Techniques 92
Field Studies 92
Conclusions 92

7. CONCLUSIONS 95
Introduction 95
Overall Scientific Conclusions 95
Specific Scientific Conclusions in Policy Context 97
Specific-Incident Criminal Investigations 97
Personnel Security Screening 99
False Negatives/Countermeasures 100
Voluntary v. Involuntary 101
Further Research 101
Concluding Comment 102

Appendixes

A. Informed Consent Forms 105
B. Results of the OTA Survey of Federal Government Polygraph Testing 107
C. Coding Form 114
D. Acronyms and Glossary 120

List of Tables

Table No. *Page*
1. Polygraph Examinations Conducted by Federal Agencies, 1982 24
2. Reviews of Field Studies of Polygraph Validity 40
3. Characteristics of Field Studies 48
4. Mean Detection Rates as a Percentage of Total in Field Studies 52
5. Outcomes of Field Studies of Validity 53
6. Outcomes of Control Question Analog Studies of Validity 63
7. Outcomes of Guilty Knowledge Analog Studies of Validity 64
8. Mean Detection Rates as a Percentage of Total in Analog Studies of Control Question Technique 65
9. Mean Detection Rates as a Percentage of Total in Analog Studies of Guilty Knowledge Test 65
B-1. Number of Polygraph Exams and Examiners 108
B-2. Purpose of Polygraph Exam 109
B-3. Length and Results of Exams 111
B-4. Long-term Confirmation of Exam Results 111
B-5. Selected Federal Agency Polygraph Research 113

List of Figures

Figure No. *Page*
1. Debate at the Federal Level 1963-83 33

Chapter 1

Summary

Chapter 1
Summary

INTRODUCTION

The primary purpose of this OTA technical memorandum is to review and evaluate current scientific evidence about the validity of polygraph testing. This memorandum responds to the February 3, 1983, letter of request from the Committee on Government Operations, U.S. House of Representatives, and the need to provide information that is relevant to congressional consideration of the polygraph aspects of the President's National Security Decision Directive-84 (NSDD-84), proposed revisions to Department of Defense (DOD) Directive 5210.48 governing the DOD polygraph program, and the recently revised administration policy on polygraph use by Federal agencies.

The OTA technical memorandum is limited to a critical review and evaluation of prior research. The memorandum does not consider, in detail, other polygraph issues such as utility, ethics, impact on employee morale and productivity, privacy, and constitutional rights. The technical memorandum, instead, focuses on the nature and application of polygraph tests, scientific controversy over polygraph testing, data from field and simulation studies, and factors that affect test validity.

FEDERAL POLYGRAPH USE

OTA found that Federal Government use of polygraph tests has more than tripled over the last 10 years, with about 23,000 examinations conducted in 1982 compared to about 7,000 exams in 1973. Current use now exceeds the previous known peak level of use (about 20,000 exams) in 1963. In all Federal agencies except the National Security Agency (NSA) and the Central Intelligence Agency (CIA), more than 90 percent of polygraph testing in 1982 was for criminal investigations. Only NSA and CIA make significant use of the polygraph for personnel security screening—preemployment, preclearance, periodic, or aperiodic—to establish initial and continuing eligibility for access to highly classified information. However, NSA accounted for almost half of all Federal polygraph examinations administered in 1982. Federal agencies at present make only limited use of the polygraph for investigation of unauthorized disclosure of sensitive or classified information—261 examinations (excluding NSA and CIA) for this purpose over the 1980-82 period.

FEDERAL POLYGRAPH POLICY CHANGES

The March 1983 draft proposed revisions to the DOD polygraph regulations (5210.48) authorize the use of polygraph tests to determine initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (Sensitive Compartmented Information and/or special access). The use of polygraph tests in determining continuing eligibility would be on an aperiodic (i.e., irregular) basis. These expanded uses of the polygraph would be part of DOD personnel security screening.

Also, the proposed revisions to DOD 5210.48 provide adverse consequences for refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access. Refusal to take an examination may, after consideration of other relevant factors, result in nonselection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position.

NSDD-84, issued by the President on March 11, 1983, authorized agencies and departments to require employees to take a polygraph examination in the course of investigations of unauthorized disclosures of classified information. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences such as administrative sanctions and denial of security clearance.

On October 19, 1983, the Department of Justice (DOJ) announced that administration policy would also permit Government-wide polygraph use in personnel security screening of employees (and applicants for positions) with access to highly classified information. The new policy provides agency heads with the authority to give polygraph examinations on a periodic or aperiodic basis to randomly selected employees with highly sensitive access, and to deny such access to employees who refuse to take a polygraph examination.

Thus, the combined effect of NSDD-84, the DOD proposals, and administration policy is to authorize substantially expanded use of the polygraph for purposes of personnel security screening and unauthorized disclosure investigations.

POLYGRAPH VALIDITY

In 1965 and again in 1976, the House Government Operations Committee concluded that there was not adequate evidence to establish the validity of the polygraph. OTA has assessed the research to determine the present state of scientific evidence.

OTA concluded that no overall measure or single, simple judgment of polygraph testing validity can be established based on available scientific evidence. Validity is the extent to which polygraph testing can accurately detect truthfulness and deception.

There are two major reasons why an overall measure of validity is not possible. First, the polygraph test is, in reality, a very complex process that is much more than the instrument. Although the instrument is essentially the same for all applications, the types of individuals tested, training of the examiner, purpose of the test, and types of questions asked, among other factors, can differ substantially. A polygraph test requires that the examiner infer deception or truthfulness based on a comparison of the person's physiological responses to various questions. For example, there are differences between the testing procedures used in criminal investigations and those used in personnel security screening. Second, the research on polygraph validity varies widely in terms of not only results, but also in the quality of research design and methodology. Thus, conclusions about scientific validity can be made only in the context of specific applications and even then must be tempered by the limitations of available research evidence.

FINDINGS

Personnel Security Screening

OTA concluded that the available research evidence does not establish the scientific validity of the polygraph test for personnel security screening. OTA was able to identify only four studies directly relevant to personnel security screening use (one by DOD). None of these studies specifically assessed validity of polygraph testing for the purposes proposed by DOD or the administration, and all had serious limitations in study design.

A 1980 survey conducted by the Director of Central Intelligence Security Committee concluded that the polygraph was the most productive of all background investigation techniques. However, this was a utility study not a validity study, and had many qualifications.

OTA recognizes that NSA and CIA believe that the polygraph is a useful screening tool. However, OTA concluded that the available research evidence does not establish the scientific validity of the polygraph for this purpose.

In addition, there is a legitimate concern that the use of polygraph tests for personnel security screening may be especially susceptible to: 1) countermeasures by persons trained to use physical movement, drugs, or other techniques to avoid detection as deceptive; and 2) false positive errors where innocent persons are incorrectly identified as deceptive.

Criminal Investigations

OTA found meaningful scientific evidence of polygraph validity only in the area of investigations of specific criminal incidents. However, OTA concluded that, even here, findings about polygraph validity must be qualified. This is because prior research has used widely varying types of questions, examiners, and examinees, among other differences. And there is, to date, no consistently used and accepted methodology for polygraph research. Also, the cases selected in field studies and situations simulated in analog studies may not be representative of most actual polygraph testing conditions. Therefore the ability to generalize from the results of prior research is limited.

OTA found a wide divergence in the results of relevant research, due in part to variations in research quality and design. Six prior research reviews showed average validity ranging from a low of 64 percent to a high of 98 percent. OTA's own review of 24 relevant studies meeting minimum acceptable scientific criteria found that, for example, correct guilty detections ranged from about 35 to 100 percent. Overall, the cumulative research evidence suggests that when used in criminal investigations, the polygraph test detects deception better than chance, but with error rates that could be considered significant.

In a typical criminal investigation, the polygraph, if used at all, is used only after prior investigation has been completed, and a prime suspect or suspects have been identified. To the extent polygraph use in unauthorized disclosure investigations would be similar, then the available research provides some evidence of polygraph testing validity. However, for so-called "dragnet" screening where a large number of people would be given polygraph tests in the investigation of unauthorized disclosures, relevant research evidence does not establish polygraph testing validity. There has been no direct scientific research on this application.

False Negatives/Countermeasures

Theoretically, polygraph testing—whether for personnel security screening or specific-incident investigations—is open to a large number of countermeasures, including physical movement or pressure, drugs, hypnosis, biofeedback, and prior experience in passing an exam. The research on countermeasures has been limited and the results—while conflicting—suggest that validity may be affected. OTA concluded that this is particularly significant to the extent that the polygraph is used and relied on for national security purposes, since even a small false negative rate (guilty person tested as nondeceptive) could have very serious consequences.

False Positives

OTA concluded that the mathematical chance of incorrect identification of innocent persons as deceptive (false positives) is highest when the polygraph is used for screening purposes. The reason is that, in screening situations, there is usually only a very small percentage of the group being screened that might be guilty. So, in a typical situation, there may be, perhaps, one person per 1,000 engaged in unauthorized activity. Therefore, even if one assumes that the polygraph is 99 percent accurate, the laws of probability indicate that one guilty person would be correctly identified as deceptive but 10 persons would be incorrectly identified (false positives). This potential problem has not been researched in field or analog studies and clearly warrants attention.

Voluntary v. Involuntary

NSDD-84, the DOD proposals, and administration policy authorize adverse consequences for refusal to take a polygraph test. Apart from the ethical and legal implications, which OTA did not address, it is generally recognized that, for the polygraph test to be accurate, the voluntary cooperation of the individual is important. Thus, OTA concluded that imposing penalties for not taking a test may create a de facto involuntary condition that increases the chances of invalid or inconclusive test results. However, no direct research on this topic was identified.

Polygraph Theory

The basic theory of polygraph testing is only partially developed and researched. The most commonly accepted theory at present is that, when the person being examined fears detection, that fear produces a measurable physiological reaction when the person responds deceptively. Thus, in this theory, the polygraph instrument is measuring the fear of detection rather than deception per se. And the examiner infers deception when the physiological response to questions about the crime or unauthorized activity is greater than the response to other questions. However, the examinee's intelligence level, state of psychological health, emotional stability, and belief in the "machine" are among the several other factors that may, at least theoretically, affect physiological responses.

A stronger theoretical base is needed for the entire range of polygraph applications. Basic polygraph research should consider the latest research from the fields of psychology, physiology, psychiatry, neuroscience, and medicine; comparison among question techniques; and measures of physiological response.

Further Research

OTA identified a need for further research on polygraph countermeasures, polygraph theory, and polygraph validity under field conditions (for both screening and criminal investigative situations). The currently planned Federal research on countermeasures appears to be inadequate. There is no known Federal research planned on polygraph theory. And the Army's current 10-year research program to develop a new, perhaps computerized, state-of-the-art polygraph instrument should be reevaluated to determine if research priorities and direction need adjustment. Finally, the planned FBI-Secret Service polygraph validity study needs an extensive independent scientific review.

CHAPTER-BY-CHAPTER OVERVIEW

The preceding discussion summarizes OTA's major findings. This section provides a brief chapter-by-chapter overview of the technical memorandum.

Chapter 2 describes the varieties of polygraph questioning techniques and a number of uses for polygraph examinations, with an emphasis on Federal Government use. The chapter describes the polygraph instrument as relatively standard, and, by itself, unable to detect truthfulness or deception. What is often referred to as "the polygraph" is actually a set of relatively complex procedures for asking questions and measuring physiological responses in order to detect deception or establish truthfulness. This chapter discusses the procedures and their common applications, and explains why different polygraph testing techniques appear to be required depending on intended uses.

The validity of polygraph examinations to detect deception has long been a controversial issue **Chapter 3** describes how the courts, State legislatures, and the executive and legislative branches of the Federal Government have viewed assessments of scientific validity as central to decisions about polygraph use. Despite many decades of discussion, no consensus has emerged about the accuracy of polygraph tests. The chapter defines scientific criteria for establishing validity and reviews previous efforts to evaluate the scientific literature on polygraph testing. Disagreement about the validity of polygraph testing in the scientific community reflects wide variations in the criteria used for inclusion of studies in prior research reviews, differences in research design and definitions of validity among specific research studies, and, perhaps most important, failure to clearly differentiate the scientific evidence in terms of the purposes for which polygraph examinations are conducted and the techniques employed.

Chapter 4 presents OTA's own analysis of polygraph field studies in order to make an independent assessment of validity. Field studies involve real-life uses of polygraph testing. With one exception, all of the available field evidence meeting minimal scientific criteria comes from cases involving specific-incident criminal investigations using the control question technique. OTA found no field studies on the validity of polygraph testing for preemployment screening or periodic screening. Overall, the studies varied in important ways with respect to, in particular, the criteria used to verify truth, and whether original examiners' decisions or blind evaluation of charts were used as the basis of comparison with ground truth. In addition, all studies had substantial problems of research design, especially with case and examiner selection. As a result, the studies may represent a highly select sample of cases. These caveats limit the confidence that can be placed in any conclusions about polygraph validity based on field research.

Chapter 5 parallels chapter 4 and presents OTA's analysis of polygraph analog studies in which field methods of polygraph examinations are used in simulated rather than real-life situations. These analog studies were conducted primarily in psychology laboratories using college students as subjects. Like the field studies, analog studies have primarily investigated the control question technique in specific-incident criminal investigations, although there are some studies of an alternative ("guilty knowledge") technique for criminal investigations and two studies of preemployment screening, one using military intelligence personnel as subjects. While using a more standardized methodology than field studies, the analog studies had other kinds of significant research design problems, and the range of error in polygraph results was greater than in field studies. The two studies of preemployment screening were of poor methodological quality, and did not adequately reflect screening for national security purposes.

Chapter 6 discusses a number of factors that may affect the accuracy of polygraph examinations. Some of these account for the variation in study results discussed previously. Examiner, subject, and setting characteristics are considered, with special attention to the use of physical, drug, and mental countermeasures that may be employed by individuals to attempt to beat the polygraph. This chapter also presents some possible priorities for further research on factors affecting polygraph validity.

Chapter 7 highlights the major conclusions and policy implications of the scientific analysis. **Appendix A** includes illustrative informed consent forms use in Federal Government polygraph examinations. **Appendix B** presents the results of OTA's survey of Federal Government polygraph use and practice. **Appendix C** includes the coding form for OTA's analysis of field and analog studies. **Appendix D** provides a list of acronyms and glossary of key terms.

CONCLUSIONS

A major reason why scientific debate over polygraph validity yields conflicting conclusions is that the validity of such a complex procedure is very difficult to assess and may vary widely from one application to another. The accuracy obtained in one situation or research study may not generalize to different situations or to different types of persons being tested. Scientifically acceptable research on polygraph testing is hard to design and conduct.

Advocates of polygraph testing argue that thousands of polygraphs have been conducted which substantiate its usefulness in criminal or screening situations. Claims of usefulness, however, are often dependent on information (e.g., confessions and admissions) obtained before or after the actual test, and on its perceived value as a deterrent.

The focus of the OTA technical memorandum is not whether the polygraph test has been useful, but whether there is a scientific basis for its use. OTA concluded that, while there is some evidence for the validity of polygraph testing as an adjunct to criminal investigations, there is very little research or scientific evidence to establish polygraph test validity in screening situations, whether they be preemployment, preclearance, periodic or aperiodic, random, or "dragnet." Substantial research beyond what is currently available or planned would have to be conducted in order to fully assess the scientific validity of the NSDD-84, DOD, and administration polygraph proposals.

Chapter 2

Varieties of Polygraph Testing and Uses

Chapter 2

Varieties of Polygraph Testing and Uses

INTRODUCTION

Polygraph examinations have been likened to psychological testing (cf. 89, 92, 101). As such, polygraph testing is best described not in the singular but, instead, as a series of tests. These tests are designed to assess truthfulness and deception in situations that range from screening job applicants to investigations of specific criminal incidents. Polygraph examiners, employed both within and outside Government agencies, use a variety of polygraph testing techniques, each of which has a somewhat different underlying logic and demonstrated validity.

The choice of polygraph technique depends primarily on the circumstances under which the polygraph is being used. The test of a subject who is suspected of a specific criminal activity typically involves application of a different polygraph technique than the examination of a prospective Government employee. Some variation in technique is also related to examiners' training, but such differences probably affect the way in which a technique is employed rather than which technique is used. A description of the instrument used in polygraph testing and an analysis of the types of test situations and polygraph techniques are presented below.

POLYGRAPH INSTRUMENT

Although there are numerous variations in testing procedures, the polygraph instrument itself is fairly standard. The polygraph measures several, usually three, physiological indicators of arousal. Changes in physiological arousal exhibited in response to a set of questions are taken to indicate deception or truthfulness. The polygraph instrument, it should be noted, is not a "lie detector" per se; i.e., it does not indicate directly whether a subject is being deceptive or truthful. There is no known physiological response that is unique to deception (108,122,123). Instead, a polygraph examiner obtains a subject's responses to a carefully structured set of questions, and based on the pattern of arousal responses, infers the subject's veracity. This assessment has been called the "diagnosis" of truthfulness or deception (139).

In actual field testing, subjects' physiological responses are measured by a three- or four-channel polygraph machine that records responses on a moving chart. Usually, three different types of physiological responses are measured. The rate and depth of respiration is measured by pneumographs strapped around the chest and the abdomen. A blood pressure cuff (sphygmomanometer) placed around the bicep is used to measure cardiovascular activity. In modern polygraph instruments, sphygmomanometer readings are electronically enhanced so as to permit lower pressure in the cuff. The electrodermal response (EDR), a measure of perspiration, requires electrodes attached to the fingertips. This has also been referred to as galvanic skin response (GSR) or skin conductance response (SCR). Each of these physiological assessments has been shown to be related to physiological arousal (36). There is some literature to suggest that one or more of the physiological channels (EDR, in particular) is most sensitive (e.g., 123). Actual field testing, however, almost always involves measurement of all three types of responses.

TYPES OF TESTING PROCEDURES

A polygraph examination normally takes anywhere from 1 to 3 hours, although shorter or longer tests may result in a variety of circumstances. The length of an examination depends on the purpose of the examination, as well as the subject's attitude and a number of other factors. Examinations may be very short because a subject "confesses" or may be lengthy when an examiner seeks to resolve an inconsistent or inconclusive pattern of responses. The examination can be divided into at least three components: pretest interview; question procedure; and post-test interview. The actual questioning aspect of the examination, which may be repeated three or four times, lasts no longer than a few minutes for each question set (limited, in some cases, because the blood pressure cuff can be inflated for only 10 to 12 minutes without causing the subject undue discomfort). Each aspect of a polygraph test is described below in detail. Unless specifically noted, generally used polygraph procedures are described. Federal Government procedures are often different and, where important such differences are noted.

The Pretest Interview

The pretest interview has been considered an indispensable component of the polygraph examination (121,139,194). The importance of the pretest is not only in its role to provide subjects with information about the examination and to inform them of their legal rights, but also in its ability to generate the psychological climate considered necessary for a valid polygraph test. An important purpose of the interview is to persuade a subject that the examination is professionally conducted and that any deception attempted "will be very obvious to the examiner" (20). Such instructions, it is thought, place truthful subjects at ease and increase anxiety in subjects who intend to be deceptive. Persuading subjects about the effectiveness of the examination should sharpen differences between deceptive and nondeceptive subjects in their reactions to questions about a particular incident.

The pretest also allows the examiner to assess the effect of special conditions or circumstances which might affect physiological responsiveness. Thus, for example, subjects are typically queried about medical problems and use of drugs that could influence autonomic responding. Such assessments are usually made without collecting "hard" data, such as blood samples.

Consent Procedures

Depending on which polygraph method is employed, as well as the subject's attitude and the situation under investigation, pretest interviews may take from 20 to 90 minutes (20,27). One aspect of the pretest interview involves obtaining the subject's consent to be examined. Consent procedures vary depending on the nature of the interview, most importantly between criminal or preemployment polygraph tests. According to Barland and Raskin (20), a typical polygraph examination conducted as part of a criminal investigation requires that the examiner advise the examinee of his or her *Miranda* rights (or rights under the Uniform Code of Military Justice). The subject is also told that the polygraph examination is voluntary. Subjects should also be informed whether or not the examination will be observed from outside the room or recorded. These disclosures are usually included in a written form which the subject is asked to sign. According to Reid and Inbau (139), criminal suspects may already have been informed of their *Miranda* rights and been asked to sign a consent form before coming to the examination room.

Applicants for employment need not be advised of their right to speak with an attorney but may, depending on local laws, need to be advised about the voluntarism of the examination. In the case of such employment-related tests, along with a provision concerning voluntary consent, subjects will be told how the results of the examination will be used. Thus, for example, they may be told that a copy of the test results will be provided to the sponsor of the exam, that the subject has a right to obtain a copy of the test results, that the subject will not be asked questions concerning such areas as political activities, union affiliations, racial or religious beliefs, or sexual activities unless these areas are specifically related to the issue under investigation (37).

Examples of consent forms used in criminal investigations by Federal agencies are shown in appendix A. The contents of Federal consent forms vary somewhat by agency, although all require that the subject "voluntarily" consent to the examination. Some agencies (e.g., Department of the Treasury (186)) indicate that the subject has the right to stop the examination at any time. Although the National Security Agency (NSA) reports that the full cooperation of the subject "is essential or the results will be inconclusive," NSA also reports (see app. B) that the polygraph examination is part of the Agency's security processing, and that failure to complete processing (which includes polygraph testing) may result in failure to be accepted for employment. As discussed more fully below (see *Current Federal Government Use*), NSA conducts polygraph examinations primarily in the context of preemployment and periodic security screening; most other agencies conduct polygraph examinations as part of specific-incident criminal investigations.

The remainder of the pretest interview also varies. In the method taught to Federal examiners at the U.S. Army Military Police School (USAMPS),* the interview focuses on questions about the subject's background: employment, family, education, health, and any previous legal problems (20). The examiner aims to learn enough to assess the subject's readiness for the examination and to prepare anxiety-provoking control questions, if they are to be used. The polygraph examiner then explains the polygraph technique to the subject and queries the subject in detail about the incident being investigated.

Another form of the pretest interview advocated by Reid (founder of the Reid College of Lie Detection) in criminal investigations makes use of a structured series of questions and deemphasizes gathering biographical data (77,139). Questions deal with matters such as the subject's suspicions about who committed the crime and the subject's feelings about the test. Questions are intended to provoke so-called "behavioral symptoms" (139) that are believed to be indicators of deception. These symptoms include evasiveness in answering, or complaints that one's physical disabilities will invalidate the recordings. When an examiner who uses the Reid method later makes an assessment of truthfulness, this information is used to supplement the data gathered from the physiological measures.

Whatever the format of the pretest interview, if control questions are to be used in the test, the last part of the interview will be used to design such questions and review them with the subject. In this phase, biographical and behavioral information collected earlier becomes essential. The information permits the examiner to tailor control questions to the individual subject. The process of designing control questions is complex and is discussed further in the section below which describes the control question technique (CQT).

Testing Procedure

Actual testing procedures have been described in detail by Barland and Raskin (20) and Reid and Inbau (139). Polygraph measuring devices, including pneumographs, a sphygmomanometer, and electrodes, are placed on the subject either during the pretest interview or at its conclusion. After the end of the pretest interview, the sphygmomanometer is inflated, and the recording of responses begins. A short period, of about 10 to 15 seconds, is used to observe initial respiratory cycles (baseline) and to allow any initial response to fade; then, the examiner asks the first question. Between each question, the examiner waits about 15 to 20 seconds until the response to the last question is finished and physiological response is closer to baseline. The examiner notes on the chart when the exam begins, when questions are asked, and when it ends. Extraneous behavior that affects the recordings may also be noted. When questions for the first chart end, the examiner deflates the cuff.

The examiner then inspects the chart and asks the subject about his or her reaction to the questions. The usual purpose for obtaining subjects' reactions is to allow refinements in the questions. The questions are reviewed again, and, when necessary, further clarified. The examiner may then administer a stimulation test, designed to improve test validity. The examiner will then continue to test and obtain two or three more charts in the same way. The examiner may use other stimulation tests between charts, and different questioning techniques (see below) to record different charts. Different questioning techniques may then be used based on information revealed by the subject. In most techniques, any new questions would be discussed with the subject before being asked. The procedure in preemployment screening or in other personnel screening tests may differ.

---

*The USAMPS provides polygraph examiner training for almost all Federal Government polygraph examiners, with the exception of CIA and NSA examiners.

Stimulation Tests

Polygraph examiners typically conduct what is known as a "stimulation"or "stim" test, designed to further convince subjects of the accuracy of the polygraph examination. Although not actually a part of the pretest, stimulation tests can be given either before the first actual set of test questions or after the first chart has been recorded. Stimulation tests are intended to reassure truthful subjects and provoke anxiety in deceptive subjects (cf. 15). Their effect should be to increase differential responsivity of deceptive and nondeceptive subjects to different questions on the examination. Some research suggests stimulation tests increase the validity of polygraph examinations (35,149).

The most common "stim" test is a "number" or "card" test. A subject is instructed to select, from a deck, a card that has a number, word, or suit on the back, or to write a number within a certain range (50,57). Sometimes, the cards are secretly marked or otherwise arranged so that the examiner is sure to know the correct answer (139). Many polygraph examiners claim this is unnecessary, however, because the technique is accurate enough without use of such deception (cf. 123), and secret markings are not employed by Federal agencies. The examiner then may repeat a range of suits, numbers or a set of words, asking the subject if each is the concealed item. The suit, number, or word that is actually the concealed item is supposed to provoke the greatest physiological response. Often, the examiner will show the subject the polygram (i.e., the actual chart recordings) to further convince subjects of the instrument's efficacy.

Types of Questions

The central element of any polygraph examination is the test of subjects' responses to a set of questions or items within questions. How these questions are structured represents the principal difference among polygraph techniques. There are four different kinds of questions or items used in polygraph testing, different combinations of questions (generally referred to as question techniques), and different applications for the various techniques. Distinctions among questions and techniques are important. Only one type of question technique in one application (CQT in criminal investigations) has been extensively researched (see chs. 4 and 5); and there are significant differences between CQT and other techniques. The range of questions, techniques, and applications is described more fully below.

Questions

The kinds of questions that are used for polygraph testing have been labeled relevant questions, control questions, irrelevant questions, and concealed information or guilty knowledge questions. Basically, relevant questions are questions about the topic under investigation (a theft, drug use, contact with foreign agents). Suspects' responses to relevant questions are of greatest interest to polygraph examiners.

Control and irrelevant questions can be grouped together as questions used for purposes of comparison to relevant questions. It is important to note, however, that the name one gives to a question may depend on the specific context in which it is used. Thus, one cannot easily give an example of a relevant question or a control question because in different situations and at different times during an examination relevant questions may be used as control questions. Likewise, irrelevant questions may become relevant, depending on a subject's response (201).

Relevant Questions

Functionally, relevant questions are questions directly related to the focus of an investigation.

In the investigation of a theft, for example, a relevant question might be "Did you steal that money?" or even more specifically, "Did you take $750 from Jones' office?" Relevant questions can be broader, however. In preemployment screening and periodic or aperiodic screening, the area of interest may be the subjects' entire background. Thus, there may be a series of relevant questions, such as "Have you ever been fired from a job?" or, "Have you stolen more than $50 in moneys in any one year from any of your employers?" (115). Intelligence agencies may ask broad questions concerning unauthorized contact with foreign intelligence agents or involvement in communist activities. Questions in an intelligence screening may also deal with areas which, potentially, may make an applicant susceptible to blackmail. It is important to note, however, that when several relevant questions relating to different issues are used, subjects are not expected to exhibit physiological responses to all of them; the relevant questions that do not evoke responses are used, after the fact, as a type of control question.

To summarize, relevant questions are questions about the topic under investigation, but topics can be very specific (Did you take $750 from Jones' office?) or cover a long period of time and a variety of acts (Have you ever stolen money from an employer? Have you ever had unauthorized contact with a foreign agent?). It is not clear what effect, if any, the breadth of a relevant question has on polygraph results, nor has there been any research done on this issue. As is discussed further in chapters 4 and 5, the preponderance of research evidence concerns the use of relevant questions to evoke reactions to specific acts.

Comparison Questions

In contrast to relevant questions, which concern issues of direct interest to the examiner, control and irrelevant questions are used for purposes of comparison. As noted above, there is no known physiological response unique to lying. Thus, a polygraph examination could not consist merely of relevant questions. If only relevant items were used, an examiner would not be able to establish the actual reason for the response. There are a number of reasons, other than fear of detection (or another hypothetically lying related reaction (19)) for a subject to become physiologically aroused during the presentation of relevant questions (48,108,136,194). Even with the addition of nonrelevant comparison items, it is necessary to run several polygraph charts using the same questions (though, perhaps in different order) to be sure that reactions are consistent. If several charts are not run, a subject's responses could be attributed to surprise, physical movement, or some reasons for concern other than a lying-related cause (203). On the other hand, the administration of several charts could theoretically just repeat the initial situation leading to the physiological response if the cause were not a random one (e.g., presence at the scene, knowledge of the incident, concern over being falsely identified). Thus, the essence of polygraph testing is the comparison of responses to the relevant questions with responses to nonrelevant questions, which have been labeled control questions and irrelevant questions.

Control Questions

Control questions, then, are used for purposes of comparison. Essentially, truthful subjects are believed by polygraph examiners to be more concerned (and, thus, more physiologically aroused) about control than relevant questions. The responses to both control and relevant questions are compared. However, control questions, like relevant questions, vary in breadth and type. One type of control question concerns what is hypothesized to be the same kind of issue that is under investigation at the time of examination. For example, a control question for "Did you take the $750 from Jones' office?" might be "Other than what you have told me [during the pretest interview], have you ever stolen anything in your life?" In an investigation of unauthorized disclosure of classified information, a control question might be "Have you ever betrayed anyone who trusted you?" Subjects innocent of the crime under investigation are presumed to be more concerned about having ever done anything of this sort (and, thus, being the "kind of person" who might have committed the crime under investigation). It is theorized that although guilty subjects will also be concerned about control questions, they will be more concerned about and thus exhibit more physiological reactions to relevant questions.

There are a number of views, however, about what distinguishes a control question from a relevant question. One distinction among control questions is whether a question is inclusive or exclusive. Inclusive control questions are questions which include the specific incident under investigation. An example of an inclusive control question in an investigation of an internal theft would be "Have you ever stolen money from an employer?" Exclusive control questions, on the other hand, cover a period of time not including the incident under investigation. An example is, "Before age 18, did you ever take anything of value?" There is some controversy over how far back in time an exclusive control question must be set for the subject to consider it psychologically separate from the incident under investigation and, thus, not a relevant question. Because inconclusive control questions may, from the suspect's perspective, include the act under investigation, some polygraphers contend that they are really relevant questions; i.e., they cannot be used for purposes of comparison. The Federal Government, for example, typically uses exclusive control questions because it views inclusive controls as relevant questions. Examiners from the private polygraph firm of John E. Reid & Associates use both inclusive and exclusive control questions.

Other kinds of nonrelevant questions other than those that cover the same kind of incident as the one under investigation, or which cover it in a different way, are also considered to be control questions. Thus, for example, "Have you ever fantasized about giving a document to a foreign agent?" is a type of control question used in some intelligence investigations. In some screening examinations, in which contact with a foreign agent is of primary concern (i.e., constitutes the relevant question), "Have you ever done anything for which you are now ashamed?" could be a control question. When a different issue than susceptibility to blackmail is under investigation, "Have you ever done anything for which you could be blackmailed?" can be used as a control question. It is noteworthy that in a different context, such as a broader screening examination, these would be considered relevant questions.

Control questions, then, are questions for which the responses are designed to be compared to responses to relevant questions. In some screening examinations, relevant questions may function as control questions after the fact. That is, if a relevant question produces a relatively mild physiological response, it may be compared to other relevant questions that produce greater response. Most often, control questions are designed to be arousing for innocent subjects (i.e., those who are not being deceptive on the relevant questions), relative at least to relevant questions. This is usually the central point of control questions, and is central to the control question technique (CQT) discussed below.

Irrelevant Questions

Another type of question used, in part, for purposes of comparison to responses to relevant questions is the so-called irrelevant question. Examples of irrelevant questions commonly used in investigations are; "Are you called [subject's name]?" or "Is today Tuesday?" Irrelevant questions are questions which are believed to have no, or very little, emotional impact on a subject. Thus, such questions can be used as an indicator of a particular subject's normal baseline level of arousal; no universal standard of physiological arousal can be applied because individuals differ markedly. Irrelevant questions are hypothesized to serve purposes other than providing a physiological baseline (139). Perhaps most important, irrelevant questions interspersed among relevant questions are hypothesized to provide a type of rest period for the subject.

Concealed Information Questions

Questions about concealed information are the fourth type of question used in polygraph testing. Unlike control and relevant questions, which ask subjects whether they have committed a crime, concealed information items aim to detect information about a crime that only a guilty subject would have. Such information might include details about the site of the crime or the means of committing it, such as the type of murder weapon used. It is hypothesized that guilty subjects will exhibit a different physiological response to the correct (relevant) detail than to the incorrect details, but that innocent subjects will respond the same to all the items. Different types of concealed information tests are described below (see *Concealed Information Tests*).

Summary

For any technique, deception is detected by comparison of suspects' physiological responses on critical or "relevant" questions or items with their responses on noncritical (irrelevant or control) items. Greater physiological responses to relevant items than to noncritical (control, irrelevant) items are assumed to be indicative of deception.

Polygraph Question Techniques

Three types of question techniques combining the four question types are described below: the relevant/irrelevant (R/I) technique, the control question technique (CQT),and concealed information techniques. Each of these test types tends to be used for particular purposes; e.g., the R/I technique is used in the great majority of preemployment screening interviews, while CQT is normally used in criminal investigations. There have been adaptations of these techniques for other uses, some of which are discussed below. Also, examiners may combine different techniques in an investigation (see, e.g., 139). In general, R/I has the broadest potential use while the concealed information techniques are the least applicable. Within each category, particularly CQT, there is considerable variability and several versions of each technique are employed.

Relevant/Irrelevant (R/I) Techniques

The R/I technique was the first standard method of polygraph questioning. The method was developed by Marston (114), a psychologist and the original proponent of polygraph examinations. An adaptation of this traditional technique is used in most of the preemployment screening conducted in the United States.

However, the R/I technique as used by the Federal Government involves somewhat different types of questions than the traditional R/I, and it must be explained separately. As currently used by Federal examiners, the R/I relies on a type of control question, and is claimed to be a version of the control question technique. The versions discussed in this section are:

1. the traditional R/I;
2. the Federal version of the R/I; and
3. the R/I as used in typical preemployment screening tests.

In a traditional R/I examination, the two types of questions used are relevant and irrelevant questions. Deceptive subjects are assumed to have a significantly greater reaction to the relevant questions than to the irrelevant questions. An underlying assumption of this technique is that nondeceptive subjects should have an equal response to all questions, because, being nondeceptive, they would not fear questions about the crime any more than irrelevant questions.

There are numerous well-recognized problems with the traditional R/I technique, at least from the perspective of psychologists who have evaluated polygraph test validity (cf. 108,126,136). First, the intent of the relevant and irrelevant questions is transparent, which means that the relevant questions are likely to be more arousing for the truthful as well as the deceptive subjects. Second, questions in the R/I technique are not usually reviewed with the subjects before the test. A larger response to the relevant question may, thus, be due to surprise or misunderstanding, as well as deception. Third, as with any question technique, reactions may be flattened by drugs or by the generally reduced responsivity of certain subjects (136). These effects are probably more difficult to detect with R/I than with other question techniques.

Because of these problems, the confidence one can place in the R/I technique is limited (136). As a consequence, the R/I technique is typically not used in the case of specific incident examinations by either public or private examiners. It is used almost exclusively with employees in nonspecific investigations. The Federal Government occasionally uses the traditional R/I and also a version of the R/I which is claimed to function as a control question test. The Federal Government version of the technique is called the general question test (GQT). Like the Reid CQT (discussed below), it uses inclusive control questions, which pertain to the subject's entire life, such that a complete answer would also include the specific incident being investigated. Thus, with a question like, "Did you ever steal anything from a place where you worked?" the theft being investigated would in actuality be part of the answer. Technically these are seen as "relevant" questions, because they are pertinent to the incident in question. Yet they are claimed to function as control questions, because they are intended to provoke a greater response in innocent subjects than questions about the misdeed provoke.

An adaptation of the R/I technique is the principal method of questioning used in preemployment and periodic or aperiodic personnel screening. Unlike the questions used with other techniques, R/I questions need not focus on one specific wrongdoing (20,108). The examiner can, thus, use the method to assess any number of issues for which the subject's veracity is to be evaluated.

In polygraph examinations used to screen employees, the polygraph examiner usually presents a series of relevant questions, with several irrelevant questions interspersed to provide a baseline. Most relevant questions ask about past behavior that might disqualify the subject from a job (e.g., employee theft, drug use, fighting on the job, incurring a large debt). Some examinations may include questions about a potential employee's background or intentions regarding a job, for example, "Did you actually graduate from college?" (201) or "Are you seeking a job with this company for any reason other than legitimate employment?" (115). Listed below is an example of questions from a preemployment screening protocol used by a commercial firm (115; also see 56,204).

Relevant questions:

Did you tell the complete truth on your job application?
Have you deliberately withheld information from your job application?
Have you ever been fired from a job?
Are you seeking a permanent position with this company?
Since the age of ( ) have you committed an undetected crime?
Since the age of ( ) have you been convicted of a crime?
During the past year, have you used marihuana (sic) more than ( ) per ( )?
Have you used any other narcotic illegally in the past ( ) years?
Have you sold marihuana (sic) or other narcotics illegally in the past ( ) years?
Have you ever stolen more than ($) worth of merchandise in any one year from any of your employers?
Have you even stolen more than ($) in moneys in any one year from any of your employers?
Have you ever used a system to cheat one of your employers?
Have you ever had your driver's license suspended or revoked?
Have you ever had any traffic citations in the past five (5) years?
Are you seeking a job with this company for any reason other than legitimate employment?
Have you deliberately lied to any of these questions?

The method used by John E. Reid & Associates employs four standard relevant questions:

In the last five years did you steal any merchandise from previous employers?
In the last five years did you steal any money from previous employers?
In the last ten years did you take part in or commit any serious crime?
Did you falsify any information on your application?

These standard questions may be modified depending on admissions made during the pretest (e.g., a revision may be, "In the last five years did you steal any merchandise other than minor office supplies?"). In addition to the standard questions a fifth relevant question (e.g., concerning the illegal purchase or sale of merchandise; use of narcotics) may be added depending on the nature of the job.

The Reid firm also uses what it regards as control questions in preemployment interviews. Control questions include, "Did you ever steal anything in your life?" and "Did you lie to any of the questions you answered during the application process for this job?" It is not clear, however, how the Reid preemployment control questions differ from the relevant questions. It seems reasonable to suppose that both truthful and nontruthful subjects (in terms of the relevant questions) may be just as concerned with the subject matter of the control questions as they are with the relevant questions. It is also not clear why employers would be less concerned with the control than with the relevant questions.

In the R/I questioning technique, a diagnosis of truthfulness or deception indicated is made by comparison of responses to each relevant question with the responses to the irrelevant questions and the remaining set of relevant questions (or in the Reid, and Army examples, control questions). Presumably, an applicant will be deceptive on no more than a few questions. These questions will provoke a greater physiological response than the others, leading to further inquiries and an eventual diagnosis (56,204).

Other types of questions are used in some screening examinations, such as questions about sexual practices or gambling. Such questions seek information about an applicant's character rather than his or her job performance and are considered by some to be unduly invasive (173). In response to this practice, ethical standards have been developed for use of the polygraph in preemployment screening (cf. 154), and some States (e.g., Illinois) prohibit their use. Preemployment polygraph examinations fall under the guidelines for emloyment interviewing of title VII of the Equal Employment Opportunity Commission, and so examiners are obliged to conduct the examinations in a way that would not discriminate on the basis of sex, race, etc. (cf. 154). One central principle of ethical standards is that relevant questions be related to the job applied for. Whether questions meet this criterion depends on the job; e.g., information about one's driving record would be important in hiring a delivery person, but not in hiring a bank teller. Screening applicants for positions involving national security apparently require questions about sexual behavior, drug use, and mental health as well as areas more directly related to national security (e.g., involvement in espionage). The range of topic areas covered in national security preemployment screening examinations by NSA is discussed below under *Current Federal Government Use*.

In so-called periodic or aperiodic checking for internal security purposes, employees are asked to submit to occasional polygraph examinations. These examinations can assess drug use, subjects' own or others' employee theft, and other matters including job satisfaction and commitment. In this type of examination, almost all of the questions are relevant questions and apparent deception (arousal) in response to any of the items is explored. Examples of the kinds of questions used in aperiodic screening in a supermarket (204), include:

Are you relatively satisfied with this job now?

Do you, as far as you know at this time, intend to stay with this employer?

Have you ever intentionally underpriced or underweighed merchandise?

Is there a particular person at your store that is responsible for damaging merchandise due to real carelessness, not caring or intentionally?

The relevant topic areas covered by NSA in a periodic screening are discussed later. Because of its use of control questions, the Federal version of R/I is discussed in the next section.

Control Question Technique (CQT)

The CQT is the most common technique used in investigations of a specific issue. The CQT was developed to deal with some of the inherent problems in the traditional R/I technique (139). Like the R/I technique, it asks relevant questions about the crime like "Did you steal the $750 from Jones' office?" As with R/I, the deceptive subject is assumed to produce a greater autonomic response to the relevant than to other questions. But CQT also adds control questions, which, as discussed briefly above, are designed to provoke a greater response in subjects who are innocent and truthful about the crime being investigated.

As discussed above, control questions are designed to be arousing for nondeceptive subjects. The questions are designed to cause innocent subjects to be doubtful and concerned about whether they have actually told the truth or told a lie. These questions usually probe for past misdeeds of the same general nature as the crime being investigated but they are transgressions that polygraphers suspect most people have "committed" or considered committing in some form (139). An example of a control question might be, "Before the age of 25, did you ever steal anything from a place you worked?" Control questions are designed to cover a long period of time, which may make the subject even more doubtful about the veracity of answers provided.

Considerable attention in the pretest interview is devoted to development of control questions (139). The process of developing control questions, reviewing them with the subject, and then refining them is designed to develop the most appropriate questions, and to convince subjects to view control questions as seriously as relevant questions. In addition, the pretest review is designed to get subjects either to be deceptive to control questions or at least to be concerned about the accuracy of their recollections (20,37,91,139). It is considered crucial to produce in the subject the right psychological set in relation to the control questions. This set is then thought to lead subjects to be more concerned about control questions than relevant questions, and so to respond more to them physiologically. This difference between response to control and relevant questions is then the basis for the diagnosis of deceptive or nondeceptive. Since the subject's psychological set is so crucial when control questions are used, differential responding to relevant or control questions (and ultimately, the validity of CQT), depends on the nature of the interaction between examiner and subject. This is true regardless of the act in question, the particular CQT method used, or the method of making assessments of truthfulness or deception. Even the validity of an entirely computerized system of scoring and diagnosis would depend on the nature of the interaction between examiner and subject. In this sense, CQT examinations, as the technology to conduct polygraph tests now stands, always require examiners to make important judgments about and interventions in their interaction with subjects.

The polygraph examiner does not tell the subject that there is a distinction between the two types of questions (control and relevant). Control questions are described as intending to determine if the subject is the "type of person" who would commit a crime such as the one being investigated (136). The examiner stresses that the subject must be able to answer the questions completely with a simple "yes" or "no" answer, that the polygraph will record any confusion, misgivings, or doubts, and that the subject should discuss any troublesome questions with the examiner (20). Thus, the situation is set up such that the subject is persuaded that the examiner wants the truth. In reality, however, the examiner wants the subject to experience considerable doubt about his or her truthfulness or even to be intentionally deceptive. According to Raskin (91), "Control questions are intentionally vague and extremely difficult to answer truthfully with an unqualified 'No'."

To produce the final version of a control question, the examiner begins by asking the subject a broad version of the question used in the pretest interview. Thus, for example, the question might be structured, "Did you ever steal anything in your life?" At this point, different polygraph examiners use slightly different methods to handle the discussion of past wrongdoing in response to the control questions asked during the pretest interview. In the USAMPS method (91), if the subject confesses to a small transgression in the past (e.g., taking home pencils from work), the examiner will dismiss it as of no consequence. For other misdeeds, the examiner will rephrase the control questions to rule them out (e.g., "Other than what we have discussed, did you ever steal anything in your life?"). The examiner will actively intervene to prevent subjects from unburdening too much of their anxiety over their past wrongs with the intention of keeping them concerned during the actual polygraph testing. Any troublesome past transgressions the subject brings up are excluded (by such phrases as "Other than what we have discussed, . . . ?") so the subject is always brought to the point at which he or she answers "No" to the control question. The control question is then ready to be used in actual testing.

The Reid method varies from the Federal method in some ways (139). If the subject does not admit to a past wrongdoing, the examiner may probe until the subject admits to one, even a crime as small as stealing pocket change from a relative during childhood. Such transgressions are then ruled out by adding the kind of exclusionary phrase discussed above (i.e., "Other than what we have discussed, . . . ?"). However, as in the USAMPS method, it is assumed at this point that the subject is either concealing other misdeeds or is worried that there are others he or she has overlooked (139). This worry has been heightened because of the examiner's emphasis on learning the truth to "ascertain" that the subject is not the kind of person that could have committed the crime referred to in the relevant questions. In addition to relevant and control questions, irrelevant questions are included during the actual interview in order to provide a baseline of physiological responsiveness.

Several versions of CQT are regularly employed and adaptations depend both on the training of the examiners and the testing situation. The Reid version can include relevant questions about several aspects of the crime (139). For example, one chart could include questions about breaking into an office, stealing a check, and then cashing it. Examiners who use Reid's CQT make a global comparison between the responses to the relevant and the responses to the control questions. They also note the subject's behavior throughout the interview (as discussed above, the Reid technique includes a series of questions in the pretest interview designed to provoke certain "behavioral symptoms" in deceptive subjects). The examiner uses the global comparison of polygraph responses supplemented by information about the behavior of the subject to make a judgment of the subject's veracity. An example of a Reid control question sequence, excluding the pretest behavior provoking items, follows (139):

1. Do they call you "Red?" (where the pretest interview had disclosed he is generally called "Red.")
2. Are you over 21 years of age? (or reference is made to some other age unquestionably but reasonably, and not ridiculously, below that of the subject.)
3. Last Saturday night did you shoot John Jones?
4. Are you in Chicago (or other city) now?
5. Did you kill John Jones?
6. Besides what you told about, did you ever steal anything else?
7. Did you ever go to school?
8. Did you steal John Jones' watch last Saturday night?
9. Do you know who shot John Jones?
10. Did you ever steal anything from a place where you worked?

In contrast, Backster's (10) zone of comparison (ZOC) technique makes a diagnosis of deceptive or truthful from a standardized numerical scoring of the charts. Each relevant question is paired with a control question. Scores are derived for each relevant question by comparing it only with the previous control question. On each physiological measure, the examiner derives a "plus" (truthful) score if the subject responds more to the control question and a "minus" (deceptive) score if the subject responds more to the relevant question. A positive score above a certain criterion level is diagnosed as truthful, a minus score below a certain level is diagnosed as deceptive, and scores in between are considered inconclusive.

A version of ZOC is used by Federal polygraph examiners. The Federal version differs from the Backster ZOC in that it may ask about several aspects of the crime in one chart. Relevant questions are asked about primary involvement (e.g., "Did you steal__________?"), secondary involvement (e.g., "Did you help steal __________?"), and so called evidence connecting (e.g., "Do you know where any of that money is now?"). In addition to relevant, control, and irrelevant questions, the Government ZOC test contains a version of the peak of tension test (see below), and "symptomatic" questions of two types. One type of symptomatic question (e.g., "Do you understand that I'm not going to ask any trick or surprise questions?") is designed to test whether the examinee trusts the examiner's word that no surprise questions will be asked. A large response is symptomatic of distrust. A second type of symptomatic question (e.g., "Is there something else you are afraid I will ask you a question about, even though I have told you I would not?") is to test whether there is some other issue the examinee is concerned about (e.g., another crime) that may be absorbing his or her arousal.

Other versions of CQT or related techniques are also used by Federal agency examiners. One, the modified general question test (MGQT), resembles the Reid CQT with the following differences: 1) only the polygraph charts are used to make determinations of truth and deception and global evaluations using inferences about behavior are dispensed with; 2) charts are numerically scored; 3) control questions exclusively concern a time and place separate from the time and place of the crime under investigation, with the intention of clearly separating responses related to the crime and the control question; and 4) the content of control questions is always related to the crime under investigation, i.e., control questions about theft are used to investigate a theft, control questions about assault are used to investigate assault, etc. Presumably, when unauthorized disclosures are at issue, control questions would concern some sort of unauthorized disclosures in the past.

To summarize, there are a number of control question techniques, the most commonly used being the Reid CQT, MGQT, and ZOC. Despite differences among them, they share the same premise and underlying rationale. Use of each of the control question procedures relies on subjects' not knowing when they are being asked the relevant and control questions. If they know which questions are more important for scoring purposes they may be able to make anticipatory responses which could invalidate their charts (see ch. 6).

Concealed Information Tests

Another polygraph questioning technique works on an entirely different premise than either CQT or R/I. Instead of detecting deception about having committed a crime per se, concealed information tests aim to detect whether a suspect has information about a crime that only a guilty subject would have or, in some cases (e.g., the actual amount of money embezzled) to detect the information itself. Such information might include details about the site of the crime or the means of committing it (e.g., the type of murder weapon used). Raskin (136) has aptly described these as "concealed information tests." Concealed information tests take two forms: the peak of tension (POT) test and the guilty knowledge test (GKT).

POT was developed by Keeler (cf. 69) and has long been used in criminal investigations. The POT test uses a set of five to nine nearly identical "yes or no" questions asking if the subject knows about a particular detail related to a crime. The detail may be a type of object used, or the color of an item. One question actually includes the relevant detail, while the others include plausible but false details of a parallel nature. The questions and the sequence in which they are asked are reviewed with the subject in the pretest interview. The subject is usually instructed to answer "no" to each question. The question with the true detail is usually presented in the middle of the sequence, so that the subject's physiological reactions will increase up to the critical question, where they will reach a peak, hence the name, and fall back down again. The card and number stimulation tests discussed above are actually examples of POT. Barland and Raskin (20) provide a hypothetical example of a POT in a criminal case:

1. Regarding the color of the stolen car, do you know it was yellow?
2. Do you know it was black?
3. Do you know it was green?
4. Do you know it was blue?
5. Do you know it was red?
6. Do you know it was white?
7. Do you know it was brown?

Occasionally, criminal investigators use the POT technique to discover and develop additional information about a case. The examiner asks the suspect about a series of details, but does not know which is actually relevant to the crime. The detail that provokes an exceptional physiological response is used as a clue in the investigation. For example, an examiner might use POT to determine the exact location where stolen goods were hidden. This kind of examination is called a searching peak of tension test (20). The searching POT technique has been used, for example, in cases in which employees are suspected of having stolen money, but there is no evidence about the extent of the theft (108). The examiner asks the employee if he has stolen money ranging from a small amount to the entire amount taken. The amount that provokes the largest response is assumed to be the amount of the total that the employee stole.

The GKT, described initially by Lykken (105, 106) works in much the same way as POT. GKT, however, often includes a larger set of questions, and the questions may be of the multiple-choice type rather than the "yes or no" type. Also, studies investigating GKT have only used the electrodermal response, while POT tests have employed standard three-channel polygraph recordings. An example of two questions from a GKT used in a laboratory study by Lykken (105) is listed below:

1. If you are the thief, you will know where the desk was located in the office in which the theft occurred. Was it (a) on the left, (b) in front, or (c) on the right?
2. The thief hid what he had stolen. Where did he hide

it? Was it (a) in the men's room, (b) on the coat rack, (c) in the office, (d) on the windowsill, or (e) in the locker?

There is a major difference, however, in the use suggested for GKT as compared to the use of the POT. POT is usually used as a supplement to a CQT, or as an aid in investigation. GKT, however, has been proposed as an alternative to control question techniques (92,107,108). Proponents argue that GKT may reduce the number of false positives, because it focuses on specific details that would be salient only to the perpetrator of a crime (108,110). Also, they claim, the validity of GKT can be substantially improved by increasing the number of questions on the test. Critics claim that it is especially susceptible to false negatives (136), that is, guilty persons not detected, and that GKT proponents do not adequately assess the consequences of false negatives.

Concealed information tests have, according to several reviewers (e.g., 108,136), other important limitations. One problem is that they may not be widely applicable. Knowledge about an incident may not differentiate between a guilty and innocent person where, for instance, a suspect is present at the scene of a crime but claims that someone else is responsible (108,136). Furthermore, concealed information tests require investigators to gather information that is not always possible to obtain, or must be disclosed to suspects in other parts of the investigation (136). In some cases, publicity about the details of a crime eliminates the possibility of a concealed information test, since the information is public knowledge (136).

POST-TEST INTERVIEW

Interspersed among test questioning and measurement of physiological responses are a number of opportunities for examiners to discuss the test with the subject. At each occasion, the examiner reviews the questions, and, depending on the responses, questions subjects about their responses. At the end of the examination, the examiner will make an assessment of whether a subject is being deceptive or nondeceptive. In some methods, e.g., Reid's (139), the assessment is a global one employing behavioral as well as polygraph data. But the USAMPS Backster's ZOC and other methods attempt to rely strictly on polygraph chart interpretation (11,20). In examinations conducted by the Federal Government, the final official determination is made after supervisory review of polygraph charts. If the subject is judged to be deceptive during the examination, the examiner will attempt to elicit a confession. Usually, this is not done directly but is couched in terms of providing the subject with an opportunity to clarify/explain the responses and differences obtained.

USES OF POLYGRAPH TESTING

As has been implied in much of the above discussion, polygraph examinations are used for a variety of purposes. The goal of all such applications of the polygraph is the detection of deception or substantiation of truthfulness. The nature of the test situation, however, leads to important differences in the way a polygraph examination is conducted. Unfortunately, the published research literature deals almost exclusively with the use of the polygraph by police and military examiners for criminal investigations. The research literature on a number of important uses of polygraph testing, such as for national security purposes and for employment screening, is extremely limited.

Current Use

The majority of uses of polygraph testing appear to be on behalf of private employers, the next greatest number are in the context of local criminal justice investigations, and the remainder are done by the Federal Government. Of greatest concern for the present analysis are the numbers and types of examinations currently conducted by agencies of the Federal Government. This section will devote most attention to such uses, although local government and private use are briefly discussed in order to place Federal use in context.

Current Federal Government Use

In order to assess the extent of polygraph use among Federal agencies, the Office of Technology Assessment (OTA) conducted a survey of Government use during May 1983. The request for information was sent to the Departments of Defense (DOD), State, Justice, Treasury, the U.S. Postal Service, and the Central Intelligence Agency (CIA), all of which were believed to employ polygraph examinations. Information was requested about the number of examinations, purposes, and results, as well as about research conducted and/or planned (see app. B). At the time of this technical memorandum, all agencies excepting CIA had provided written responses to the request for information about the number and type of polygraph examinations that have been administered.

CIA declined to respond because of the classified nature of the information. However, some data about CIA's use for background investigations were reported in a 1980 study (165). The number of polygraph examinations are summarized in table 1. Table 1 indicates that Federal agencies reported administering a total of 22,597 polygraph examinations in fiscal year 1982. As shown in appendix B, about half of these were in the context of criminal investigations. Polygraph examinations are also reported to be used for intelligence and counterintelligence investigations (some (NSA) at aperiodic intervals), and preemployment screening. The largest single number of polygraph examinations conducted in 1982 were conducted by NSA, a component of DOD, primarily for preemployment screening. These numbers can be compared to previous surveys conducted in 1963, when Federal agencies, excluding NSA and CIA, conducted 19,796 polygraph examinations, and 1973, when 6,946 examinations (including 3,081 by NSA) were conducted.

Table 1.—Polygraph Examinations Conducted by Federal Agencies, 1982[a]

Agency[b]	Total
Department of Defense:	
Army Criminal Investigation Command	3,731
Army, Intelligence Command	279
Navy	1,337
Air Force	3,019
Marines	263
National Security Agency	9,672
Department of Justice:	
Federal Bureau of Investigation	2,463
Drug Enforcement Agency	211
Department of the Treasury:	
Secret Service	714
Bureau of Alcohol Tobacco and Firearms	256
U.S. Postal Service	652
Central Intelligence Agency	n.a.[c]
Totals	22,597

[a] Data were also reported for fiscal years 1980, 1981, and, in some cases year to date 1983. See app. B for complete report.
[b] The U.S. Customs Service (Department of the Treasury), Department of Health and Human Services, and Tennessee Valley Authority conduct a limited but unknown number of polygraph examinations.
[c] Classified or partly classified.

SOURCE: Office of Technology Assessment.

As shown in appendix B, NSA reports that it uses primarily the R/I technique. NSA reports that counterintelligence-type screening examinations—i.e., tests given to NSA (or affiliated) personnel who already have access to classified information—would have relevant questions on the topics of involvement in espionage or sabotage against the United States; knowledge of others involved in espionage or sabotage against the United States; involvement in giving or selling classified materials to unauthorized persons; knowledge of others giving or selling classified material to unauthorized persons; and unauthorized contact with representatives of a foreign government (187). Examinations that are given to applicants for employment and contractors who are applying for access to Sensitive Compartmented Information (SCI) consist of questions about the topics covered in counterintelligence-type aperiodic screenings (phrased as "Do you plan to commit. . . ?") as well as questions about a broader range of issues: involvement in communist, fascist, or terrorist activity; commission of a serious crime; involvement in adult homosexual activity; involvement with illegal drugs or narcotics; deliberate falsification of security processing forms; treatment for a serious nervous or mental problem (187). According to NSA, the scope of specific issue examinations is limited to questions that are relevant to the issue to be resolved. Presumably, specific issue examinations would be conducted using the control question technique.

Current DOD regulations also allow the use of polygraph examinations to investigate situations in which credible derogatory information about an individual with clearance is provided to officials. The frequency of this type of investigation, however, was not reported. Prior to the President's National Security Decision Directive of March 11, 1983, use of the polygraph in personnel investigations of competitive service applicants and appointees to competitive service positions was limited to executive agencies with highly sensitive intelligence or counterintelligence missions affecting the national security (e.g., a mission approaching the sensitivity of that of CIA; see 188). Approval to use the polygraph could be granted for only 1-year periods. Refusal to consent to a polygraph could not be made a part of an applicant or appointee's personnel file. See chapter 3 for a description of proposed changes in Federal use of polygraph testing.

Non-Federal Government Use

Outside the Federal Government, polygraph examinations are administered as part of criminal investigations, as well as preemployment screening and periodic screening of employees for purposes of controlling internal crime and recommending promotions. Less frequent uses include examinations in such situations as paternity investigations and workers' compensation cases. It has been estimated that over a million polygraph examinations are given a year (107), 300,000 of them for employment purposes alone (128).

Both private and police polygraphers use polygraph examinations in the process of criminal investigations (see 136). In some cases (most typically, rape and kidnapping cases, but also, for example, investigations of improper or illegal conduct by public officials (177)), witnesses and victims whose veracity is in doubt are asked to take a polygraph examination. Suspects who claim innocence may be asked by their defense attorneys or the prosecution to support their claim by taking a polygraph examination. In such cases, prosecutors and defense attorneys may make informal agreements to drop the charges if the polygraph examination indicates no deception. Or, the prosecution and the defense may formally stipulate that if deception is indicated, results of the polygraph examination will be admissible at trial. In some cases (New Mexico, Massachusetts, and the 9th Federal Circuit Court of Appeals (8,136, 156,157)) polygraph evidence has been admitted over objection. Polygraph evidence is also used occasionally in postconviction proceedings such as sentencing and motions for a new trial (136). In polygraph examinations as part of criminal investigations, some version of the control question technique is typically used.

The use of the polygraph examination by employers is reported to be widespread (144). Although it is illegal to ask employees to take an examination in 19 States and the District of Columbia, it is legal to do so in 31 States (8,156,157). Polygraph examinations are used most commonly in commercial banking, investment banking, and retail operations. In such settings, both risk of theft and fraud are high and, in addition, employee turnover is high. The use of polygraph examinations is also recommended to employers as a check before making promotion decisions (204).

CONCLUSIONS

What is often referred to as "the polygraph" is actually a set of relatively complex procedures for asking questions and measuring physiological responses in order to detect deception or establish truth. Polygraph testing is employed for a variety of uses, ranging from ascertaining the guilt of a criminal suspect to assessing the honesty of a prospective employee. Because different polygraph procedures are required depending on intended use, it is necessary to consider validity by polygraph technique and situation. In subsequent chapters, such a variegated analysis is presented and the scientific and policy contexts are more fully described.

Chapter 3

Controversy Over Polygraph Testing Validity

Chapter 3

Controversy Over Polygraph Testing Validity

INTRODUCTION

The validity of polygraph examinations to detect deception has long been a controversial issue (cf. 108,136,194,195). Since development of polygraph techniques almost 80 years ago, their use both within and outside the Federal Government has been the focus of numerous judicial opinions and, as well, legislative and executive branch debate. Polygraph examinations have been advocated as a way to ascertain guilt of criminal suspects, to exculpate innocent suspects, to protect national security, and to maintain employee honesty. Polygraph examinations have, at the same time, been criticized for providing inaccurate and misleading information, for failing to detect security risks (167), for interfering with the rights of private citizens (128), and for lowering employees' morale. At the center of controversy over the use of polygraph examinations is the question of its validity: does a polygraph examination actually identify truthful and nontruthful individuals?

Recent interest in polygraph examinations and their validity stems from efforts to broaden Federal Government use. The Department of Defense (DOD), in late 1982, drafted revisions to existing regulations (5210.48). DOD proposed expansion of the use of polygraph tests for preemployment screening and periodic or aperiodic testing of employees who have access to highly classified information. Currently, only the National Security Agency (NSA) and the Central Intelligence Agency (CIA) are able to use polygraph tests in this way. Expanded use of polygraph testing in all Federal agencies was made explicit in a Presidential National Security Decision Directive (Mar. 11, 1983, NSDD-84). In part, the directive requires agencies and departments which handle classified information to revise existing regulations to permit use of polygraph examinations as part of internal investigations of unauthorized disclosure of classified information. Prior to the directive, investigations of unauthorized disclosures had to be referred to the Department of Justice (DOJ). Employees who refuse to submit to a polygraph examination could, if NSDD-84 is implemented, be subject to adverse consequences. In October 1983, DOJ announced that administration policy would also permit Government-wide polygraph use in personnel security screening of employees (and applicants for positions) with access to highly classified information.

Proposals to expand use of polygraph examinations to maintain national security have renewed the debate about the appropriateness of various polygraph techniques and their ability to detect deception. In order to provide a context for the present evaluation of scientific evidence on the validity of polygraph testing, previous assessments of accuracy of polygraph testing are reviewed in this chapter. Legal precedents regarding polygraph testing and congressional hearings on its use, both within and outside of Government, are briefly considered. The chapter also describes scientific criteria for establishing validity and reviews other efforts to evaluate the scientific literature on testing.

JUDICIAL REVIEWS

When courts have been called on to resolve disputes concerned with use of polygraph examinations, they have had to consider both the technique's validity and whether its use, however valid, interfaces with other values that the law seeks to protect. The varying decisions reached by State appellate courts and Federal circuits (see 8) may in large measure reflect varying beliefs about the validity of polygraph examinations. Indeed, for many years, the leading case on the admissibility of novel scientific evidence (*Frye* v. *United States* (58)) was a case about the admissibility of polygraph evidence, and the opinion centered on the question of validity. The issue of how a court is to decide the question of any scientific technique's validity has brought the *Frye* test into question in recent years and makes salient the problem of establishing judicial standards for assessing validity (60).

Polygraph Findings as Evidence

The *Frye* case involved a 19-year-old defendant convicted of robbery and murder. Prior to his trial, a well-known psychologist and one of the originators of polygraph testing, Dr. William Marston, administered a "systolic blood pressure test" to detect deception (e.g., 114). Dr. Marston determined, on the basis of this test, that Frye was truthful when he denied involvement in the robbery and murder. The trial judge, however, refused to permit Dr. Marston to either testify about the examination or conduct a reexamination using the blood pressure test in court.

Frye appealed his conviction on the grounds that relevant exculpatory evidence had not been admitted. The appeals court, however, concurred with the initial trial court judgment. The court reasoned that the systolic blood pressure deception test was validated only by "experimental" evidence and was not based on a "well-recognized scientific principle or discovery." The decision stated that, "while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the things from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. Just when a scientific principle crosses the line between experimental and demonstrable is difficult to define."

Ironically, Frye's conviction was later reversed when another man confessed to the crime, thereby providing Frye with more convincing corroboration of his denials of guilt. This did not settle the case, however, and recent discussion of the facts of the case indicate that Frye was, indeed, guilty. The crude polygraph examination conducted by Marston, thus, appears to have yielded an inaccurate conclusion.

The *Frye* test is still used as precedent in most Federal courts. Subsequent opinions (in areas other than the polygraph) have tried to better define that line between "experimental" and "demonstrative" stages of a scientific innovation. For example, the court in *United States* v. *Stifel* (190) held that "neither newness nor lack of absolute certainty in a test suffices to render it inadmissible in court." In a second case, *United States* v. *Brown* (189), the court also seemed to be concerned with validity: "The fate of a defendant in a criminal prosecution should not hang on his ability to successfully rebut scientific evidence which bears an 'aura of special reliability and trustworthiness,' although, in reality the witness is testifying on the basis of an unproved hypothesis in an isolated experiment which has yet to gain general acceptance in its field." The *Frye* test has been held to be too high a hurdle by some trial courts, which have replaced it with the test for admissibility of expert testimony generally: "testimony by a witness as to matters which are beyond the ken of the layman will be admissible if relevant and the witness is qualified to give an opinion as to the specialized area of knowledge" (190).

A closely related question for the courts has been who should determine whether some procedure has gained general acceptance in its field. Some have held that the courts must look to the judgment of the scientific community (e.g., 191). In other decisions, the court refused to "surrender to scientists the responsibility for determining the reliability of (scientific) evidence," and that "a determination of reliability cannot rest on a process of 'counting (scientific) noses.' "

Saks and Van Duizend (145) concluded that whichever set of tests is employed, the courts are in a weak position to assess validity directly or to count scientific noses. The result has been: 1) general deference by the courts to the judgments of scientific communities; and 2) "numerous incongruities . . . where less reliable scientific and technological information is admitted but the admission of demonstrably more reliable techniques is delayed until the requisite consensus has formed" (145; see, also, 60).

When the courts examine polygraph testing, they are faced with a series of dilemmas. To which "particular field" of expertise can the courts turn: physiology, psychology, polygraphy? If they look to the data themselves, what are they to make of it? As the present report suggests, validity assessment involves a complex situation and technique-specific answer. Even if a final, single accuracy rate could be established, how should a court use it. How accurate must a diagnostic or predictive technique be to be deemed valid for evidentiary purposes? Regularly admitted psychiatric evidence is widely recognized (including by the U.S. Supreme Court, see *Addington* v. *Texas,* (2)) as having accuracy rates comparable to flipping coins (e.g., 55,208). In *Barefoot* v. *Estelle* (13) the Supreme Court acknowledged that psychiatric predictions of dangerousness and violent behavior do not exceed an accuracy level of 33 percent (see 118). Yet, this evidence was held admissible in *Barefoot* and sufficiently valid to uphold a decision to execute a convicted person.

In summary, then, the courts have found themselves disagreeing on methods to establish validity for purposes of admissibility of evidence, where the critical focus of such judgment should rest. In addition, courts are inconsistent about what decision to make on the basis of judicial findings of fact regarding the validity of a diagnostic or predictive device.

Laws Regulating Polygraphs in Employment Settings

As described in chapter 2, screening employees is the most frequent application of polygraph testing. Many employers argue that use of polygraph testing for preemployment screening, periodic checking, and to resolve actual thefts is necessary. Internal crime has been estimated to cost private industry up to $10 billion annually (see 172), and polygraph testing is regarded as a cost-effective tool. Employers argue that screening applicants, and periodic checking of employees, are the most efficient ways to control pilferage, embezzlement, poaching, and other forms of theft. The need for polygraph testing is felt particularly in industries which have high risk of theft and fraud (e.g., commercial banks), high turnover (supermarkets, other retail operations), or both.

According to Ansley (8), the use of private polygraph testing is limited by statute in 18 States plus the District of Columbia. Most of these laws seek to protect employees from being requested, required, demanded, or subjected to polygraph examinations by their employers. Employers are reported to be able to find ways around these laws. For example, employers may tell the employee that they suspect them of theft, but that if the employee can find a way to demonstrate innocence, the employer will not discharge the employee. In addition to polygraph validity, other polygraph-related concerns include issues of voluntariness, invasions of privacy, being compelled to inform on other employees, inhibiting union activity, and the polygraph as a cover for racism and sexism. This list does not exhaust concerns that have been expressed.

A survey of 143 private firms by Belt and Holden (25), regarding their use of polygraph testing, yielded a number of interesting findings. Twenty percent of respondents reported using polygraph examinations for preemployment screening, periodic surveys, and investigations of specific onsite crimes. It is interesting that of reasons given for using or not using polygraph tests, users ranked moral or ethical considerations last and efficiency first; nonusers, however, ranked validity and reliability second in importance, cost third, and the availability of qualified operators fourth in importance. The survey found a positive relationship between a State having a licensing requirement for polygraphers and employers' use of polygraph testing. According to Ansley (8), 25 States have licensing requirements for polygraphers; licensing is optional in one State.

Although there is testimony that use of polygraph testing reduces employee crime (172), no formal cost-benefit analyses appear to have been conducted. In addition, there is no research on the predictive validity of polygraph results (72,144). Although employee issues are critical to proposed Government uses of polygraph testing, few data are available on Government employees (see chs. 4 and 5).

One additional area of controversy has concerned employee rights and employer-employee relationships. The general matter of invasion of privacy is particularly pertinent in preemployment screening and periodic checking. In preemployment screening, the range of questions that may be asked has been subject to particularly heavy criticism. Questions have been reported to include items concerning union activity, sexual preference, and family problems (169); and, in addition, willingness to make a commitment to the job (144), and whether the respondent has ever been tempted to steal (71). During periodic checking, respondents are sometimes asked not only about their own possible improper behavior (e.g., underringing in supermarkets), but also about their level of job satisfaction, intention to remain with the employer, and activities of their fellow employees (204). There is some concern about whether prejudices of the polygraph examiner based on racial, ethnic, and gender stereotypes bias employees' responses (144). These assertions do not appear to have been researched. And no related claims under Title VII of the Civil Rights Acts have been upheld.

One argument against the use of polygraph examinations in the employment situation is that it destroys the trust relationship between employers and employees, and creates employee dissatisfaction. However, the few employee surveys that have been conducted have not supported this argument. Apparently, five studies have examined whether the use of the polygraph causes private sector employees to be dissatisfied. In one study (144), 96 percent of applicants were willing to take a polygraph examination to get a job, 86 percent of the applicants thought the preemployment examination was fair, and 88 percent were willing to take it routinely as a condition of employment. A problem with the study was that applicants were surveyed immediately after taking the polygraph examination so they may have thought their responses were part of the screening process. In the one known survey of Federal employees, the Air Force (183a) surveyed individuals who had volunteered to participate in a pilot project on the use of the polygraph for counterintelligence/security examinations. About 99 percent of the respondents felt that the examinaton was fair, and were willing to take an examination for counterintelligence purposes.

FEDERAL DEBATE OVER POLYGRAPH VALIDITY

Concern about and debate over Federal Government use of the polygraph have emerged at several points during the past 20 years. As shown in figure 1, the history is essentially one of legislative concern triggered by some executive branch proposal or action regarding polygraph testing. The questions raised by Congress have included constitutional and ethical as well as validity issues. However, the scientific validity and reliability of polygraph testing has been and is a central congressional concern. This chapter briefly describes the history of Federal Government involvement with the issue of polygraph validity.

The 1960's

Congressional interest first intensified in 1963 when controversy developed over an executive branch proposal to use lie detectors to find the source of unauthorized disclosures of sensitive or classified information, sometimes known as "leaks" (192). The then chairman of the House Committee on Government Operations asked the Foreign Operations and Government Information subcommittee to study the Federal Government's use of polygraphs. The study found that, excluding the National Security Agency and Central Intelligence Agency (for which information was classified), Federal agencies had conducted 19,796 polygraph examinations in 1963. In 1964, the subcommittee held hearings and received testimony from private polygraphers, researchers, and Federal officials. In a 1965 report (167), the House Committee on Government Operations concluded that there was no scientific evidence to support the theory of the polygraph, and that the research

Figure 1.—Debate at the Federal Level 1963-83

	1963-64	1965	1966	1967-74	1974	1975	1976	1977-78	1979	1980	1982	1983
Executive branch activity	Proposal to use polygraph to investigate leak	DOD Directive 5210.48 DOD research program Civil Service regulations				DOD issued revisions to Directive 5210.48				Director of Central Intelligence study Investigative Scope and Adjudicative Procedures Among Intelligence: Personnel Security Survey	DOD Select Panel Review of the DOD Personnel Security Program Deputy Secretary of Defense issues memos instituting aperiodic polygraph exams for those with SCI clearance DOD prepares proposed revisions to Directive 5210.48	President issues NSDD-84 DOD prepares further draft of revised regulations DOJ announces administration policy on polygraph testing
Legislative activity on Federal agency use	Study and hearings by House Government Operations Subcommittee on Foreign Operations and Government Information	Report issued by House Government Operations Committee		Bills introduced which would restrict or ban Federal Government use of polygraph	Report Privacy, Polygraphs and Employment issued by Senate Judiciary Subcommittee on Constitutional Rights Study and hearings by House Government Operations Subcommittee on Foreign Operations and Government Information		Report issued by House Government Operations Committee	S.1845 introduced: would bar Federal agencies and private employers from requiring or requesting Hearings held on S.1845 by Senate Judiciary Subcommittee on the Constitution	Hearings and report by House Permanent Select Committee on Intelligence Subcommittee on Oversight		Hearings by House Judiciary Subcommittee on Civil and Constitutional Rights on DOD proposed revisions	Hearings by House Post Office and Civil Service Subcommittee on Civil Service and House Judiciary Subcommittee on Civil and Constitutional Rights on NSDD-84 Congress passes amendment to DOD appropriations bill prohibiting changes in DOD polygraph regulations until Apr. 15, 1984 Hearings by Senate Committee on Government Affairs on NSDD-84
Legislative activity on private employer use				Some bills would ban use of polygraphs by private employers	Report referred to private as well as Federal employer			S.1845 introduced Hearings Privacy Protection Study Commission issues report	Hearings held by House Education and Labor Subcommittee on Labor-Management Relations			HR2403 introduced: would ban use of polygraph by private employers

evidence as to its accuracy was inadequate. The committee recommended that further research be conducted and training for polygraph examiners be upgraded, and that the President establish an interagency committee to study and work out solutions to problems posed by Federal Government use of polygraphs.

Later in 1965, an interagency polygraph committee of representatives from DOD, CIA, DOJ, Bureau of the Budget (now Office of Management and Budget), Office of Science and Technology (now the Office of Science and Technology Policy), and other executive agencies was established. The interagency committee concluded that: 1) there was insufficient scientific evidence concerning the validity and reliability of polygraph testing; and 2) the use of the polygraph constituted an invasion of privacy of the individual being interrogated. The committee recommended that the "use of the polygraph in the executive branch should be generally prohibited, and permitted only in special national security operations and in certain specified criminal cases" (166). The recommendations made at that time concerning personnel screening were promulgated as Civil Service regulations on regulating the use of polygraphs in personnel investigations of competitive service applicants and appointees to competitive service positions (ch. 736, app. D, of the Federal Personnel Manual). According to these regulations, which are still in effect, only executive agencies with highly sensitive intelligence or counterintelligence missions directly affecting the national security such as "a mission approaching the sensitivity of that of the Central Intelligence Agency" are permitted to use the polygraph for employment screening and personnel investigations of applicants for and appointees to competitive service positions. All other uses of a polygraph to screen applicants for and appointees to competitive positions are forbidden.

The regulations also set forth steps for determining whether agencies met the criteria of having a highly sensitive mission, and stipulated that approval to use the polygraph would be granted only for 1-year periods. Agencies intending to use the polygraph for personnel screening were required to prepare regulations and directives meeting certain minimum standards. The minimum standards included directives concerning the specific purposes for which the polygraph may be used, and directives that a person to be examined must be informed as far in advance as possible of the intent to use the polygraph and of the fact that refusal to consent to a polygraph examination will not be made a part of the person's personnel file.

Also in response to the House Government Operations Committee's 1965 report, DOD proposed, and in part undertook, an extensive polygraph research program. And in July 1965, DOD issued directive 5210.48 (177) to regulate the conduct of polygraph examinations and improve selection, training, and supervision of its polygraph operators. Some of the results of the DOD research program were later reported in a scientific journal (29), but other reliability and validity studies proposed were never carried out (183).

Between 1967 and 1973 a number of bills were introduced which would have either limited the questions that could have been asked or banned altogether polygraph use by Federal agencies (170). None of these bills was enacted.

The 1970's

Ten years after the 1964 hearings, this same House Government Operations subcommittee conducted another review of polygraph use by Federal agencies (169). In 1974 hearings, the subcommittee found that the use of polygraphs in the Federal Government had declined substantially since 1963. In fiscal year 1973, a total of 6,946 examinations were conducted, including 3,081 by NSA. This compared to 19,796 in 1963, excluding NSA and CIA. The subcommittee also found that there was not much additional research on polygraph validity. The only federally funded studies conducted had been those reported by the DOD Joint Services Group (183), and these studies were considered by DOD to be inadequate for determining the validity and reliability of Federal polygraph testing.

In a 1976 report based partly on the 1974 hearings, the House Government Operations Committee concluded that "the nature of research undertaken, both federally and privately funded, and the results therefrom, have done little to persuade the committee that polygraphs . . . have demonstrated either their validity or reliability in differentiating between truth and deception, other than possibly in a laboratory situation" (171). The 1976 report concurred with the 1965 report that "There is no 'lie detector' " (171). Because of the polygraph's "unproven technical validity" and the suggestion that the "inherent chilling effect on individuals subjected to such examination clearly outweighs any purported benefit to the investigative function of the agency," the Committee recommended a complete ban on the use of polygraphs by all Federal Government agencies for all purposes. However, 13 committee members dissented, asserting both that the hearings had been held during an entirely different Congress, and participated in by an entirely different group of Members, and that, while testimony at the hearings represented a wide diversity of views, no witness had urged prohibition of the polygraph for all purposes. The dissenters urged adoption of the recommendations originally proposed and voted on by the members who had participated in the hearings. These recommendations would have, in part, prohibited the use of polygraphs in all cases except "1) those clearly involving the Nation's security, and 2) those in which agencies can demonstrate in compelling terms their need for use of such devices for their law enforcement purposes, and that such uses would not violate the fifth amendment or any other provision of the Constitution."

The concern with scientific validity and its implications for the Federal Government's use of polygraph testing arose again in 1979 at hearings held on preemployment security clearance procedures by the House Permanent Select Committee on Intelligence, Subcommittee on Oversight (175). The subcommittee found that there had been insufficient research on the accuracy of the polygraph technique in screening job applicants and that "gaps in the statistics kept by the intelligence services do not make it possible to make the clear judgment that the polygraph is unique and indispensable" (173). The Director of Central Intelligence (DCI) was urged to conduct a study to validate the accuracy of the polygraph for preemployment screening. DCI did conduct a study in 1980 to examine the utility of polygraph tests, but it was not a validity study (165).

As shown in figure 1, in addition to interest in Federal use of polygraphs, Congress has shown interest in the use of polygraph examinations by private employers, in part because of constitutional and privacy issues (see, e.g., 169,172,173; the Privacy Protection Study Commission Report (128) mandated by Public Law 93-579; and several laws introduced since 1967). Various congressional committees have questioned the validity of polygraph testing in a private employment context, in particular as a condition for employment. Nevertheless, attempts to enact Federal legislation regulating the use of polygraph examinations by private employers and/or the Federal Government have not been successful.

The 1980's

In the recent past, the executive branch has again taken initiatives concerning the Federal use of polygraph testing. In April 1982, a DOD select panel reviewed the DOD personnel security program (180) and expressed dissatisfaction because of inconsistency in polygraph use across component programs (as did the U.S. Congress (173)), and the lack of reinvestigations. The panel observed that military personnel, unlike civilians, were appointed to NSA and allowed access to Sensitive Compartmented Information (SCI) without undergoing a polygraph examination. In addition, personnel could continue to get clearances throughout their careers without ever being subject to reexamination. The DOD panel recommended a broadened application of the polygraph for security screening purposes, and selective use of counterintelligence scope polygraph examinations during periodic reinvestigations. The panel noted that the recommended expanded use of the polygraph would require changes in DOD Directive 5210.48.

On August 6, 1982, the Office of the Deputy Secretary of Defense (39) issued a memorandum requiring employees with SCI access to agree to submit to polygraph examinations on an aperiodic basis, and revised DOD Directive 5210.48 accordingly. Later in 1982 and again in early and mid-1983, further revisions to DOD Directive 5210.48

were drafted (181). In 1983, the President issued a National Security Decision Directive (NSDD-84) also authorizing broader use of the polygraph. Congress responded to these developments by conducting several sets of hearings, by requesting OTA and General Accounting Office studies, and by passing an amendment to the DOD appropriations authorization bill (S.675) putting a moratorium until April 15, 1984, on any revisions to DOD Directive 5210.48 retroactive to August 5, 1982. On October 19, 1983, DOJ announced a new administration polygraph policy that would permit further expansion in polygraph use. The DOD draft revisions, NSDD-84, and administration polygraph policy are discussed in more detail below.

Draft Revisions to DOD 5210.48

The draft revisions to the DOD polygraph regulations have gone through several iterations. For the purposes of this validity study, a primary proposed revision (as of the March 1983 draft) is to authorize the use of the polygraph for determining initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (SCI and/or special access). The use of the polygraph in determining continuing eligibility would be on an aperiodic (i.e., irregular) basis (181).

Also, the proposed revisions provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may, after consideration of all other relevant factors, result in adverse consequences for the individual. Adverse consequences are defined to include nonselection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position.

Technically, these expanded uses of the polygraph are considered to be part of personnel security investigations. Use of the polygraph within DOD is already authorized under the existing 1975 version of 5210.48 for various criminal, counterintelligence, and intelligence purposes.

A detailed review of the proposed changes is beyond the scope of this technical memorandum.

NSDD-84

On March 11, 1983, the President issued a National Security Decision Directive intended, according to DOJ officials, to help safeguard against unlawful disclosure of properly classified information. One provision of NSDD-84 requires that persons with authorized access to classified information sign a nondisclosure agreement, and that persons with access to SCI must also agree to prepublication review. These provisions are outside the scope of this memorandum, as is a full analysis of NSDD-84.

With respect to the polygraph, NSDD-84 in effect authorizes agencies and departments to require employees to take a polygraph examination in the course of internal investigations of unauthorized disclosures of classified examinations. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences. NSDD-84 permits administrative sanctions, including denial of security clearance, to be applied even when a person is not subject to a criminal investigation (184).

Administration Polygraph Policy

On October 19, 1983, DOJ announced a comprehensive administration policy on Federal agency polygraph use. The policy authorizes polygraph testing:

1. as a condition of initial or continuing employment with or assignment to agencies with highly sensitive responsibilities directly affecting national security;
2. as a condition of access to highly sensitive categories of classified information;
3. to investigate serious criminal cases; and
4. to investigate serious administrative misconduct cases including unauthorized disclosure of classified information (185a).

The policy in essence authorizes use of the polygraph on a Government-wide basis for the expanded polygraph uses proposed by DOD. Thus, for example, the policy provides agency heads with the authority to give polygraph examinations on a periodic or aperiodic basis to randomly selected employees with access to highly sensitive information, and to deny such access to employees refusing to take a polygraph exam.

SCIENTIFIC VALIDITY AND POLYGRAPH RESEARCH REVIEWS

Thus, recent polygraph policy actions have renewed interest in and debate over the scientific validity of the polygraph. Reviews of scientific literature form the principal means to cumulate research findings and are especially important in order to assess the validity of polygraph testing. Single research studies, no matter how well conducted, cannot answer global questions about validity and must be considered in relation to other evidence. Both because research evidence about polygraph testing has rapidly increased, especially within the last 10 years, and because there have been disagreements about the nature of evidence about polygraph testing, there have been a number of such reviews. These reviews are important, because they are frequently cited in both legal and legislative considerations and because they serve to shape future research.

Underlying each of the reviews is the application of a set of criteria, only sometimes made explicit, regarding the validity of individual studies and their implications for overall assessments of polygraph testing accuracy. As introduction to the scientific reviews, the nature of these criteria is described. The reviews, themselves, are then summarized and a preliminary analysis of discrepancies among reviews is presented. More detailed analysis of individual validity studies is provided in chapters 4 and 5.

Definitions of Scientific Validity

Validity

The validity of polygraph testing means, in nontechnical terms, accuracy of the test in detecting deception and truthfulness. The problem of assessing polygraph validity is especially difficult, not only because polygraph tests take a number of forms, but also because validity has different dimensions and can be measured in a number of ways. There are, as a result, a number of different forms of validity associated with polygraph examinations depending on the type of polygraph test as well as on its use (e.g., employee screening v. investigation of a criminal suspect). These difficulties underlie, in part, the failure to have developed assessments of polygraph validity that are accepted by the scientific community.

In order to make explicit the criteria for validity used in this assessment, below are described several dimensions of validity and how they are assessed. This description is based both on standards for psychological/psychometric tests (cf. 3,5) and criteria to evaluate research designs (cf. 41,147). Although criteria for validity can be described objectively, it should be noted that it is essentially a qualitative judgment as to whether (or, to what extent) a given criterion is met. In addition, assessments of the "preponderance" of evidence necessary in order to assess the overall validity of polygraph testing are similarly subjective. In chapters 4 and 5, a systematic analysis of available research is attempted, although it should be recognized that there are a number of ways to conduct such evaluations, each of which may yield a somewhat different outcome.

Reliability

Assessment of any test's validity is based on the assumption that the test consistently measures the same properties. This consistency, known as reliability, is usually the degree to which a test yields repeatable results (i.e., the extent to which the same individual retested is scored similarly).

Reliability also refers to consistency across examiners/scorers. A reliable polygraph test should yield equivalent outcomes when subjects are retested and, as well, be scored similarly by individuals other than the initial examiner. For example, if a polygraph examiner reviewed a set of charts and concluded that a subject was deceptive, any other polygraph examiner should be able to review the same charts and conclude that deception was indicated. This illustrates interrater-reliability. Such reliability might be affected by the amount and type of training of examiners.

The present study focuses primarily on validity because if a testing procedure is not measuring what it purports to measure (validity), it matters little that it can measure the same thing again and again. Examiners who consistently agree that they are seeing "deception" may in fact be measuring anxiety or some other form of arousal. Reliability is, however, a necessary condition for validity to be established. A test that is valid will, necessarily, be reliable.

Construct Validity

Construct validity refers, in broad terms, to whether a test adequately measures the underlying trait it is designed to assess. A polygraph test is designed to detect deception. It is therefore important to clearly define the construct of deception, and distinguish it from other concepts such as guilt.

To measure construct validity, it is necessary to both describe the construct and show its relation to a conceptual framework. Construct validation, thus, requires that a test be based on some theory or conceptual model. Since different types of polygraph tests have different theoretical bases (see ch. 2), there are multiple forms of construct validity for the polygraph. Construct validity is established by various means. Most importantly, based on theoretical predictions of how items should interrelate or how other tests should intercorrelate, actual evidence (e.g., scores from similar tests) is examined. If no such predictions are possible, it is impossible to establish construct validity.

Criterion Validity

Although from a theoretical point of view construct validity is most important, from a practical point-of-view, criterion validity is the central component of a validity analysis. This aspect of validity refers, in the case of polygraph examinations, to the relationship between test outcomes and a criterion of ground truth. In this respect, criterion validity is what is meant by test accuracy. In the absence of construct validity evidence, however, it is difficult to determine to what extent criterion validity data can be generalized. In some situations, it is not clear which aspects of a test are responsible for accuracy, and what factors cause a test to be inaccurate.

Research Design

The above validity criteria are those which are typically assessed in considering evidence about the usefulness of a test. A related set of validity crtieria are also used to evaluate the validity of any single study design. These research design criteria include, most importantly, internal and external validity (cf. 41,147).

Internal validity refers to the degree to which a study has controlled for extraneous variables which may be related to the study outcome. External validity refers to the established generalizability of a study to particular subject populations and settings. Internal validity in the case of a study of polygraph testing is usually enhanced by the presence of control groups. Typically, such conditions of an experiment permit analysis of variables such as different question formats. In most field studies, internal validity is difficult to establish since the investigation cannot control or, in many cases, have definitive knowledge about whether a subject is guilty or innocent.

External validity is simply the nature of the subjects and settings tested. The broader the population examined and the type of setting investigated, the wider that study's results can be generalized. In a parallel way, the more similar the research situation to the "real life" situation, the greater a study's external validity. Evidence about external validity is developed both from investigations that test a broad range of subjects and situations and from investigations that identify subject and setting interactions with polygraph test outcomes. The broader the population examined and the type of setting or the more similar it is to the situation for which one wants to use a test or a theoretical construct, the greater a study's external validity.

False Positives and Negatives

With any test, the possibility exists of false positives and negatives. False positives are decisions that individuals are being deceptive when they are providing truthful responses. Their charts are scored as showing a "deceptive" reaction for some other reason. False negatives are decisions that individuals are not being deceptive when in fact they are being deceptive. There are a number of reasons why such false outcomes might be obtained and, in part, they depend on the criteria (e.g., amount of physiological change) used to indicate deception or truthfulness.

The rate of false positives or negatives is sometimes difficult to establish because, in research studies, a number of criteria for deception/nondeception may be applied. Thus, for example, in studies which employ numerical scoring for polygraph charts, depending on the scoring system (e.g., cutoff points), different diagnoses will be made. The rate of false positives and negatives may also depend on the examiner's perception of the "base rate" of guilt/innocence.

In some cases, the examiner will deal mostly with deceptive subjects (e.g., in certain criminal investigation contexts) and, thus, may be predisposed to make false positive diagnoses. In other settings (e.g., some personnel screenings), an examiner may test only a small number of deceptive subjects and, then, may be predisposed to false negative decisions. Regardless of rates, assessment of conditions that contribute to either type of error is a focus of the research literature.

Reviews of Polygraph Validity

Since at least 1973, a number of polygraph researchers and psychologists interested in physiological detection of deception have reviewed available scientific literature to assess the validity and reliability of polygraph testing. Most such reviews focus on studies of criterion validity, although a growing number of investigations deal with construct validity. The most important difference among these criterion studies has to do with whether they are conducted in actual field situations or in "analog" situations.

Field Studies

For purposes of this technical memorandum, field studies are those studies or "naturally" occurring polygraph test situations; i.e., studies in which the researcher does not exercise experimental control over the situation in which the crime or other event occurred. Not exercising experimental control means that the researcher does not systematically assign people to conditions of, for example, guilt or innocence. We refer here to "field" studies but others (e.g., 7) use the terminology "real" cases (v. "laboratory"). Abrams (1) differentiates between the laboratory and "actual criminal cases."

In polygraph field studies, polygraph examiners' decisions are compared against some post hoc determination of whether suspects are guilty or innocent; i.e., "ground truth." These post hoc determinations may, in different studies, consist of confessions by the presumably guilty party, decisions by a panel of attorneys or judges assembled specifically for a particular study who base their decisions on investigative files excluding references to polygraph decisions, judicial outcomes (dismissals, acquittals, convictions), as well as other criteria. The fact that determinations of guilt or innocence are made post hoc makes drawing conclusions from field studies difficult (126). In real life situations, truth is seldom available (62).

Attempts to use confessions, panel judgments, judicial outcomes, and other criteria as indicators of truth have their own problems. Individuals may confess to crimes which they did not commit (108). In addition, individuals are sometimes falsely convicted (34). Panel decisions may be generalizable only to cases in which sufficient investigative information is available to make a decision without the addition of polygraph testing. One can never be certain that the panel decision is indeed correct, and the panel and the polygraph examiner may have been exposed to the same prior information (62). Thus, while field studies provide the most direct evidence about polygraph test validity, they have been criticized because they do not adequately meet the standards of "ground truth" to establish criterion validity.

Comparison of Reviews

A number of independent reviews (listed in table 2) of the field evidence on polygraph testing were assessed in order to determine reasons for differences among reviews. The reviews differ in a number of respects. In part, reviewers' conclusions differ because they include different kinds of studies and even different studies (despite, in several cases, having had the same studies available to them). In addition, some reviews differentiate between accuracy in detecting deceptive v. nondeceptive subjects, emphasizing the problems of false positives and false negatives; others aggregated the overall accuracy rates across both groups of subjects. Finally, there are differences

Table 2.—Reviews of Field Studies of Polygraph Validity

Study[b]/N[e]	Year	Q[c]	D[d]	Abrams (1973)	Horvath (1976)	Raskin & Podlesny (1979)[a]	Lykken (1981)	Ben-Shakhar (1982)	Ansley (1983)[i]
				Reported "hit rate" (%)[f]					
Bitterman and Marcuse (N = 2Ex, 81S)	1947	R/I		[100]	NR	NR	EX[g]	NR	[100]
Ben-Ishai (N = 1E, 100C)	1962 unpublished	C	A	NR[h]	NR	NR	NR	NR	[97.0][i]
Ben-Ishai (N = 1E, 10C)	1962 unpublished	C	A	NR	NR	NR	NR	NR	[100]
Bersh (N = 4E, 216C)	1969	C	G	NR		NR		85.6	
			I					89.3	
			A		[74.6-92.4][j]		[75-92][j]		[96.2][i]
Horvath and Reid	1971	C	G	[79.1-91.4][k]	EX[l]	75-89[k]	EX[l]	85.0	NR
			I			89-94[k]		90.5	
Hunter and Ash	1973	C	G	—	EX[l]	88	EX[l]	87.1	NR
			I			86		85.7	
Wicklander and Hunter	1975	C	G	—	EX[l]	95	EX[l]	90.0	NR
			I			93		86.7	
Slowick and Buckley	1975	C	G		EX[l]	85	EX[l]	83.8	NR
			I			93		90.5	
Raskin (3 studies)	1976[n]	C	G	—	—	93-100[m]	NR	NR	NR
			I			69-95[m]			
Barland and Raskin (N = 102C)	1976	C	G	—		NR	98	83.0	NR
			I				45	29.4	
			A		[78.7-89.7][o]		[72]		
Horvath	1977	C	G	—	EX[l]	NR	77	77.1	NR
			I				51	51.1	
Davidson	1979	C		—	--	—	NR	NR	EX[l]
Edwards (N = 80E, 959C)	1981 unpublished		A	—	—	—	NR	NR	[98.3][p]
Elaad (N = 184C)	1976 unpublished		A	—	—	NR	NR	NR	[98.3][q]
Peters (N = 172C)	1982 unpublished		A	—	—	—	—	—	[95.1][i]
Widacki (N = 38C)	1982	Z	A	—	—	—	—	—	[95.8][i]
Author's conclusion (if any)			G			90		83-84[r]	
			I			89		76-81[r]	
			A	98††	NR		64-72[s]		97.6

---

[a] Not technically a review. Original review (126) relied on lab evidence. When criticized by Lykken, authors responded in part by citing field studies.

[b] Presented in chronological order.

[c] Q = question type (C = control question; R/I = relevant/irrelevant; Z = zone of comparison).

[d] D = correct decision or rate of agreement (i.e., hit rate) for guilty (deceptive) (G) or innocent (nondeceptive) (I) subjects.
A = average hit rate (in cases where G and I are not reported separately).

[e] N = number. Ex = number of examiners; E = number of evaluators; C = number of cases evaluated; S = number of subjects when actual subjects are tested in the study.

[f] Different studies use different criteria for validity. Thus, the hit rate can be either polygraphers' agreement with ground truth (e.g., Bitterman and Marcuse), blind evaluators' agreement with polygraph examinations verified by confessions (e.g., Horvath and Reid, Hunter and Ash, Wicklander and Hunter, Slowick and Buckley, Horvath), or polygraph examinations validated by a panel of judges using other criteria (e.g., Bersh, Barland, and Raskin) or actual judicial outcome (Widacki). Researchers have not agreed on what constitutes an appropriate criterion for validity in the field.

[g] Excluded by Lykken as not useful evidence because of base rate problem. the examiners in the study assumed only one person was guilty. Therefore, they would have had a 99-percent accuracy rate even if they had called everyone innocent.

[h] NR = not reported by the reviewed indicated.

[i] It is important to note Ansley's method of calculating overall accuracy. When there are inconclusives (or disagreements between the original examiners and the evaluator), Ansley divides the percent inconclusive by 2 and adds one-half the percentage to the accuracy percent reported by the original researcher.

[j] The 74.6 percent figure is for majority decision; 92.4 percent for unanimous decision.

[k] The lower rate is the accuracy rate for inexperienced polygraphers; the higher rate is the accuracy rate for experienced polygraphers.

[l] Excluded because they are seen as reliability studies. Horvath reasons that they used confessions as a criterion and because confessions are not independent of polygraphers' examinations "their usefulness as a criterion measure for estimating validity is limited." Lykken excludes them because both examiners and evaluators were Reid polygraphers.

[m] Lower rate is for nonnumerical evaluation; higher rate for numerical evaluations.

[n] Focus of study was to test the "friendly polygrapher" hypothesis (see Raskin, Barland, and Podlesny, 1978).

[o] 78.7 percent = agreement between simple majority on panel and polygraph decision; 86.7 percent = agreement between 4/5 majority on panel and polygraph; 89.7 percent = agreement between judicial outcome and polygraph (where judiciary not aware of polygraph results).

[p] Of 2,433 polygraph examinations in the original pool, only 959 (39.4 percent) could be verified as to truthfulness or not; of these 959, 943 were found to be correct. The 98.3 percent figure reported by Ansley represents 943/959 and does not take into account the entire pool of examinations.

[q] Actually, of 1984 cases that were confirmed, 10 were inconclusive and 6 were errors. The percent accuracy given in Ansley's review excludes the inconclusive results and adds half the error rate to the accuracy rate. If one appropriately excludes both and inconclusives and the error rate, the accuracy rate is at most 87 percent.

[r] The lower rate is for studies using confessions criteria; higher rate is for studies using panel of experts criteria. Figures are rounded.

[s] Lie control test only.

†Calculations based on Lykken's conclusion that 6 out of 11 innocent subjects were judged deceptive.

††Abrams uses many other studies to come to this conclusion. Many are anecdotal. Other researchers (e.g., Horvath) say they should not be relied on.

in the way accuracy rates were calculated, in particular, how inconclusives are handled. Each of these differences has important implications for the conclusions developed by the reviews.

Several reviews (1,81) conducted 5 to 10 years ago reported relatively positive conclusions based on an evaluation of the scientific literature.

Abrams (1) in 1973 reviewed reports of the polygraph's accuracy dating from 1917, including anecdotal as well as experimental data. He calculated approximate estimates of overall accuracy from this data, noting, however, that "it is almost meaningless to total and average these findings because of the great discrepancy in experimental paradigms and the instruments employed." He reported that in studies with complete verification of ground truth, diagnoses were 100 percent correct. In other field studies prior to 1963 Abrams calculated an accuracy rate of 98 percent. In laboratory experiments prior to 1963, Abrams estimated the average accuracy rate of 81 percent. Averaging the results of the reports between 1963 and 1973, Abrams estimate of laboratory and field research accuracy was 83 and 98 percent, respectively. Horvath's (6) review in 1976 used somewhat more stringent criteria in selecting data than did Abrams. His review does not include an overall average accuracy rate calculated across studies.

The early positive views of the polygraph's worth have recently been challenged by Lykken (108) and, to some extent, by Ben-Shakhar, et al. (28). Lykken in 1981 challenged the theoretical assumptions of the most prevalent question technique, the control question technique (CQT), and asserted that an average 50-percent false positive rate supported his theoretical challenge. Lykken, however, continues to believe that particular polygraph techniques are useful (i.e., the detection of guilt by measuring physiological arousal) and offers the use of the guilty knowledge technique as a way to increase overall validity. Adoption of Lykken's suggestion would preclude the use of the polygraph for preemployment testing and periodic checking.

Ben-Shakhar, et al.'s (28), analysis also limited their assessment of the polygraph to CQT. Their 1982 assessment of existing polygraph field research indicated that polygraph testing was 83 to 84 percent accurate for guilty suspects and 76 to 81 percent accurate for innocent suspects. As a result, Ben-Shakhar, et al., concluded that examiners tend to value detection of guilty suspects highly, even at the risk of falsely classifying innocent suspects; their conclusion concurs with Lykken's. Ben-Shakhar, et al., in conductng their review, employ a utility theory approach based on Bayes' theorem. They predict dramatically different utility rates based on different base rate assumptions.

Although these recent reviews, by authors who are not professional polygraphers, cast doubt on the validity of at least the most common polygraph technique, a more recent review by Ansley (7) comes to the most positive conclusions since those of Abrams. Ansley's 1983 review is an important review because it represents the views of NSA's chief polygraph examiner. (NSA conducts the largest number of polygraph examinations of any Federal agency.) As shown in table 2, Ansley concludes that field research shows a 97.2-percent validity rate and laboratory research a 93.2-percent validity rate. Based on these validity calculations as well as separate calculations for reliability and utility, Ansley concludes that the polygraph is "clearly an excellent adjunct to the selection process."

Unfortunately, for the most part, polygraph reviews contained in table 2 do not explicitly state their study selection criteria (see 63). The result is that a number of different studies have been included in various reviews, each of which presents different problems for interpretations of validity. The kinds of studies include reports of single criminal investigations in which the actual solution to the crime is the criterion for validity; studies in which "blind" polygraph interpreters compare their polygraph chart evaluations to "ground truth" as established by confession; and studies in which the judgment of legal professionals, actual judicial outcome, or in one case, the judgment of a single psychologist, is used to establish ground truth.

Some reviews do specify criteria for exclusion. Lykken, for example, does not include studies of single criminal investigations. Abrams, on the other hand, includes in his review a number of such studies (e.g., 30,103). Lykken's reasoning was that in single criminal investigations, the examiner has a large chance of being accurate (depending on the number of suspects) merely by calling everyone innocent. The fact that other reviewers do not include Bitterman and Marcuse, and other such reports, implies that they accept Lykken's evaluation of the usefulness of such studies as indicators of validity. It is possible that results of such reports could be useful in assessing polygraph screening of large numbers of individuals in specific incident cases, such as might be the case in unauthorized disclosure investigations. However, additional factors limit the external validity of Bitterman and Marcuse and other such studies. In Bitterman and Marcuse, for example, the investigators were psychology professors apparently conducting their first polygraph tests, and they did not use accepted polygraph procedures or instruments. There are no recent systematic studies of specific incident investigations involving a large number of suspects.

There is strong disagreement among reviewers about whether another group of studies should be included as indicators of validity. These studies were conducted with records selected from the files of the John E. Reid & Associates polygraph firm. A group of cases was used which the authors considered to be "verified" by confession of the guilty suspect (in most cases they were also verified by some form of corroboration (37)). The polygraph charts in these cases are then reinterpreted by a group of polygraphers who are "blind" to (i.e., do not know) the suspect's guilt or innocence. The degree of agreement of the "blind" evaluators to verify guilt or innocence is the test of validity. Two reviewers (Horvath, Lykken) explicitly excluded the group of studies conducted based on Reid files. Horvath excluded them because they used confessions as a criterion (confessions not being independent of the polygraph examinations), and Lykken because both examiners and "blind" evaluators were polygraphers from the same firm. His claim was that the studies were, thus, "merely demonstrations that Reid's examiners score charts in a similar way" (108) and so were estimates of reliability rather than validity. However, reviews by Raskin and Podlesny (138) and Ben-Shakhar, et al. (27), each use all four Reid studies to assess validity.

Conclusions about the validity of the polygraph may depend on whether the reviewer attends to the average accuracy rate or to the accuracy for guilty and innocent subjects separately. The conclusions of all decision statistics contributes to the ability to make an accurate assessment of polygraph testing validity, particularly in view of the concern over both high false positive and high false negative decisions. If, for example, the innocent correct rate is 80 percent but the remaining 20 percent consists of inaccurately calling innocent subjects guilty, a different policy conclusion may be drawn than if the remaining 20 percent consists of "inconclusives" or of false negatives. In some cases (e.g., preemployment screening), inaccurately designating nondeceptive people as deceptive may have worse consequences for the employee than inaccurately deciding that deceptive individuals are nondeceptive. In some cases (e.g., a heinous crime by a potential repeat offender, infiltration by a foreign agent), a false negative may have serious consequences.

In only two reviews (Ben-Shakhar, Lykken) are summary percentages provided in terms of the percent accurately detected for both guilty and innocent; in other reviews, these figures are presented as the average percent of accurate detections. In some cases, the percent inaccurately "detected" as nondeceptive (when they were really deceptive) or deceptive (when they were really nondeceptive) as well as percent inconclusives were also reported by reviewers. But for purposes of clarity these have been omitted from table 2.

Another reason reviews differ about the results of the same studies is the fact that they make different decisions about the base rate of subjects or cases that are included. If, for example, a panel cannot make a decision about 30 percent of the cases (e.g., 22), some reviewers will omit the number of nonagreements from the number included in the accuracy rate and base accuracy percentages on only the remaining cases. This accounts for the difference between Horvath and Ben-Shakhar, et al., analyses of the Barland and Raskin results. In other studies (and reviews of those studies, e.g., Ansley, Abrams) inconclusive polygraph results are excluded from the analysis. This has the effect of inflating the accuracy rates.

Apart from the different base rates on which most of the reviewers calculated accuracy rates (see above), one source of different accuracy rates applies uniquely to Ansley (7). In any case in which there is not 100-percent accuracy, the Ansley review computes validity by dividing the difference between the accuracy rate and 100 percent (the so-called error rate) in half and adds half of the difference to the accuracy rate. Ansley uses this procedure on the grounds that on the basis of chance, errors were probably half in favor of the panel (or other criterion measure) and half in favor of the examiners. For example, in the Bersh study, half of the difference between the typically reported 92.4-percent rate and 100 percent is 7.6 which Ansley divides in half, leaving a validity rate of 96.2 and an error rate of 3.8 percent. The same method is used for the Peters, Elaad, and Widacki studies, for which the preadjustment validity rates are 90.2, 96.6, and 91.6 percent, respectively. Each of these studies, particularly Elaad (see ch. 4), have other problems of interpretation as well.

CONCLUSIONS

Central to legal, legislative, and scientific assessment of polygraph tests are their validity. Yet, despite many decades of judicial, legislative, and scientific discussion, no consensus has emerged about the accuracy of polygraph tests. One explanation is that scientific criteria for validity deal with a number of dimensions and that the criteria vary widely among specific research studies. In order to assess overall polygraph examination validity, it will be necessary to examine details of each of the relevant studies. Such analysis is presented in chapters 4 and 5.

Another explanation is that polygraph testing has been viewed as a single technique. Thus, despite testimony (e.g., 137) which urged differential consideration of polygraphs used in, for example, employment screening and criminal investigations, the scientific evidence for particular purposes has not been differentiated. As is demonstrated by the analysis of scientific literature (here and in chs. 5 and 6), in assessing validity it is necessary to separate clearly the purposes for which polygraph examinations are conducted and the types of techniques employed.

Chapter 4

Review and Analysis of Polygraph Field Studies

Chapter 4

Review and Analysis of Polygraph Field Studies

INTRODUCTION

As noted in the discussion of previous scientific reviews of polygraph validity, considerable disagreement exists among reviewers as to which field studies and what kinds of evidence constitute acceptable tests of validity. This chapter presents the results of a systematic analysis of existing field studies of polygraph testing in order to make an independent assessment of validity. Field studies investigate actual polygraph examinations and constitute the most direct evidence for polygraph test validity (27). Both quantitative and qualitative techniques are utilized in order to make an overall assessment of existing evidence (63,125,142).

The goal of this analysis is to synthesize available research. Almost all of the available field evidence comes from cases involving specific-incident criminal investigations using the control question technique (CQT). This is an important limitation. Because a systematic review helps to identify this kind of problem, researchers and policymakers have a better basis on which to determine what, if any, additional studies are necessary. Also, the analysis aids understanding of which question techniques, test purposes, question designs, and scoring techniques have been studied and which may require further research. The analysis is designed to address many of the problems associated with qualitative or "literary" reviews of the research literature previously discussed. In particular, the analysis makes explicit the criteria used for both study selection and data analysis (63,125,142).

STUDY SELECTION

Studies were considered field studies of validity if their sample consisted of actual instances of polygraph examinations conducted by professional polygraph examiners, used field-tested polygraph techniques, and used some independent criterion to assess actual guilt or innocence. Although ground truth can probably never be known in an absolute sense, studies can be considered studies of validity only if they included some adequately described and systematically determined criterion of "truth" (e.g., panel decision, judicial outcome, confession). Studies in which judgments of one set of polygraphers are correlated with anothers' with no independent criterion of guilt or innocence are, in effect, reliability studies. Such studies have been excluded from the primary analysis reported here. Reports of unsystematically collected cases from police agencies and other organizations, in which the criteria for verification are unclear or unsystematic, have also been excluded.

The population of field studies considered for the present analysis was, in general, taken from those studies referred to in existing reviews of the scientific literature (see ch. 3). In addition, researchers active in the field of polygraph research were contacted and asked to supply the names and publication information of any additional recent studies. A bibliography provided by the American Polygraph Association (9) was also searched for references to field studies of validity. The 10 studies finally included (and listed in table 3) in the analysis are: Barland and Raskin (22), Bersh (29), Davidson (47), Horvath (83), Horvath and Reid (84), Hunter and Ash (85), Kleinmuntz and Szucko (92), Raskin (133), Slowick and Buckley (155), and Wicklander and Hunter (205). The fol-

Table 3.—Characteristics of Field Studies

	Type of validity affected		
Study	Criterion	Basis of Examiner Decision	Types of cases[a]
Bersh	Panel of legal professionals' assessment of investigative files	Original examiners' decisions	Criminal investigations/military personnel
Barland and Raskin	Panel of legal professionals' assessment of investigative files	Original examiners' decisions	Sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes [b]
Barland and Raskin[c]	Panel	Blind evaluation	Sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes
Barland and Raskin	Judicial outcome	Original decision	Sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes
Barland and Raskin[c]	Judicial outcome	Blind evaluation	Sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes
Raskin	Confession	Blind evaluation	Sex crimes, drug crimes, crimes of violence, crimes of financial gain, other crimes
Horvath and Reid	Confession	Blind evaluation	Theft, sexual misconduct, sabotage, bribery, criminal damage to property
Hunter and Ash	Confession	Blind evaluation	Theft, official misconduct, brutality, sexual assaults, homicide
Slowick and Buckley	Confession	Blind evaluation	Theft, industrial sabotage, drug abuse, rape
Wicklander and Hunter	Confession	Blind evaluation[d]	Homicide, sexual assault, theft, official misconduct
Horvath	Confession	Blind evaluation	Crimes against persons, crimes against property
Davidson	Confession	Blind evaluation	Crimes against property/military personnel
Kleinmuntz and Szucko	Confession	Blind evaluation	Theft

[a] All studies use some version of control question technique.
[b] Only 77 of 92 cases were analyzed as to type of crime.
[c] Not included in the analysis for reasons discussed in the text.
[d] Wicklander and Hunter also included an evaluation in which evaluators were given additional case material.

lowing sections briefly describe the studies excluded from the analysis and the kinds of studies included in the analysis.

Studies Excluded

Not all studies referred to as field studies or actual criminal investigations by other reviewers are included in the present analysis. A comparison of studies shown in table 2 and the 10 studies included in the present analysis indicates that eight studies included by one or another of the reviewers are not included. The excluded studies are Bitterman and Marcuse (30), Ben-Ishai (26), two analyses reported in Raskin (133), Edwards (52), Elaad and Schahar (54), Peters (124), and Widacki (206). One study, Kleinmutz and Szucko (92), not included by various reviewers (because of its recent publication) has been included here. In addition, a number of studies included by Abrams (1), not shown in table 2, are also excluded from the present analysis. Many of the studies Abrams cited are excluded by later reviewers (e.g., Horvath (81)) because they are not actual validity studies (and did not use external criteria of "guilt/innocence," e.g., MacNitt (113)), they did not use appropriate polygraphic instrumentation (e.g., Summers; see Abrams (1)), or did not use testing procedures common today (e.g., Lyon (111)). Other studies used by Abrams, but excluded from the present analysis, were unverified self-reports published in popular magazines (e.g., McEvoy (116)), or surveys of attitudes towards validity of the polygraph (e.g., Cureton (44)).

The Bitterman and Marcuse (30) study was excluded because, as pointed out by Lykken (108) and Horvath (81), among others, studies of single crimes for which there is only one possible guilty person raises the probability of accurate deception, regardless of method used, to a level too high for the study to provide valid information. To give an extreme example, if there is one guilty suspect among 100 examined, making an a priori decision to call them all innocent yields a 99-percent accuracy rate. In addition, Bitterman and Marcuse did not meet present criteria for field studies because the polygraphers were not professional examiners (they were psychology professors who had read books and articles about the polygraph technique), and they did not use field-tested measures of physiological response.

Ben-Ishai's (26) paper reports on two studies, both of which were excluded. One consisted of blind evaluations by Ben-Ishai of 10 polygraph charts. It is more accurately described as a study of reliability. The other used a single psychologist's (Ben-Ishai's) judgments of guilt or innocence based on investigative files as the criterion by which to judge polygraph accuracy. It is difficult to justify use of the judgment of a single psychologist as an adequate criterion of ground truth. Likewise, the information used to establish ground truth for the Elaad, Peters, and Widacki reports is not systematically collected and is inadequately described. These studies are more accurately described as a set of anecdotal reports. They use samples of cases collected from police files which are described as having been verified, sometimes by judicial outcome (Widacki), in others by confession (Elaad), and in the Edwards study, by "independent means."

A final set of studies excluded are two of the three studies by Raskin (133). One analysis was directed primarily at an assessment of whether polygraph examinations are more favorable to defendants when conducted by polygraph examiners chosen by defense attorneys than when they are conducted by examiners chosen by prosecutors (the so-called "friendly polygrapher" hypothesis). The purpose of the second analysis was to discover the source of decision errors; these findings are discussed in chapter 6. The Raskin study included in the present analysis (133) was conducted with only the 16 cases from Barland and Raskin's (22) sample able to be verified by confession.

Studies Included

The field studies included are listed in table 3 in terms of the criterion used, the type of initial examiner decision, and the types of cases selected. These characteristics of studies relate to criterion, construct, and external validity, respectively.

The criterion dimension refers to the operationalization of ground truth used in a study. In one type of validity study, polygraphers' original decisions are compared against a criterion of ground truth established by a panel of experts (e.g., lawyers and judges). The panel makes their judgment on the basis of information in an investigative file, from which polygraph results are excluded. In another type of field study, a second set of examiners evaluates charts taken from a file. In most cases, the evaluation is "blind;" i.e., the examiner/evaluator does not know the original examiner's decision, the disposition of the case, nor any other information about the subject. In this situation, the original decisions have been verified by confession of the guilty party. Verification by confession is used as the ground truth criterion. In the third, and the least common type of field study, original examiners' decisions (the construct validity component) are judged against guilt or innocence established by judicial outcome, which is the ground truth criterion.

Researchers disagree about whether blind evaluations of polygraph charts or the decisions of the original examiners constitute true tests of polygraph validity. Whether one uses examiner decisions or physiological recordings depends on whether one is testing examiner decisionmaking or physiological arousal in response to certain questions. Blind evaluations of charts are probably less useful as research evidence because, in the typical examination situation, the decision as to suspects' deception is made by the original examiner and not by a blind evaluator. Even when examinations are subject to review (e.g., quality control procedures used by the Department of Defense (DOD)), final decisions are still based on review of all information. Although a blind analysis is the first task of the quality control office, such quality control reviews do not fully control for the impact of a variety of factors, such as interpersonal expectancy effects which would still be reflected in the original polygraph charts. Interpersonal expectancy effects (141) refer to the possibility that an examiner's preexamination decision concerning guilt or innocence affects construction of examination questions or the psychological state of the suspect. Either of these could affect a suspect's physiological responses. Therefore, in studies for which results of both original examinations and blind evaluations were included, as in Barland and Raskin (22), the present analysis uses results of the original examinations instead of those for blind evaluations. It should be noted, however, that in these cases it is difficult to determine to what extent the decisions are based on the charts and to what extent they are based on interaction with the suspect (see 27,92).

Operationalizations of ground truth (the criterion component of validity) are also problematic. Studies using panel decisions have been referred to as the only valid field research on the validity of examiners' decisions (81), yet there is no way to know whether panel decisions based on investigative files are, in fact, correct. Raskin (136) notes some of the problems with using judicial outcomes and other criminal justice system resolutions (dismissals, guilty pleas) as criteria for validity. Cases may be dismissed for lack of sufficient evidence rather than actual innocence. If a jury acquits a defendant, it is not possible to determine the extent to which the jury felt that the defendant was actually innocent or whether they felt that there was not enough evidence to meet the standard of "guilty beyond a reasonable doubt." Many guilty pleas are actually confessions of guilty to (lesser) crimes; as Raskin notes, it is difficult to interpret the meaning of such pleadings in regard to guilt on the original charge. The result is that, using criminal justice system outcomes, polygraph examinations may appear to have a high number of false positives (in the case of acquittals), or false negatives (in the case of dismissals).

The use of confessions, the most frequently used criterion of ground truth, is problematic in three ways:

1. confessions, themselves, are not always valid;
2. if the confession occurs prior to or during a polygraph examination, it cannot be considered an independent measure of guilt; and
3. those who confess may be a select sample of subjects, as discussed further below.

In addition to the above problems, studies differ in the adequacy of their research design. The most serious problems concern sampling. In most reported studies, neither cases, examiners nor evaluators were selected randomly. In some studies (e.g., 22,84), the cases of only one examiner are sampled. Nonrandom selection leaves open the possibility that the studies are not investigating "polygraph testing" in general, but instead only a subgroup of practitioners or testing techniques. When random sampling is used (as in Bersh (29)), high rejection rates of cases selected for analysis create other sample bias problems.

Some sample selectivity of unknown magnitude and importance occurs when confessions are used as a criterion. Studies using confessions may be using only a select sample of examinations. The magnitude of this problem is illustrated by the fact that in the sample of 92 cases obtained by Barland (22,133) only 16 were able to be verified by confession (132).

To summarize, because of problems in operationalizing important components of validity, none of the field studies of validity can be taken by itself as an indication of polygraph testing validity. In addition, because of the different operationalizations of construct and criterion validity and variations in research design, the studies are not strictly comparable with each other. These studies, however, constitute the most direct evidence for validity currently available and are analyzed as a group in order to assess the current state of knowledge about polygraph testing.

CODING

In order to conduct the present analysis, each field study was coded for a number of variables which had either been referred to as important factors in previous reviews of the literature, or which were deemed relevant to the various components of validity described in chapter 3. If the needed information was not available from the studies as published, the study author(s) were contacted and asked to supply the information. Appendix C lists the coding categories including relevant validity components (panel decision or judicial outcomes; confession), as well as design information (sample selection, attrition rate, examiner/evaluators' knowledge of base rate of guilt). All codings were made by two reviewers and each instance of disagreement over coding was resolved before analysis.

Data were coded directly from information provided within the study report or from information directly provided by the authors, with the exception of one variable. The exception was the coding category "objectivity of ratings," which required that the coder make a judgment from high objectivity to low objectivity. Scoring was judged high if some actual standardized measurement (e.g., using a ruler) was taken of the physiological recordings on the polygraph charts. A rating of medium was given if numerical scores were assigned to subjective assessments of suspects' guilt or innocence (see, e.g., 22,92), low if ratings of deceptive or nondeceptive were based on global assessments of charts only, and very low if decisions were based on charts plus other available information (in particular, observation and interaction with the subject). Objectivity ratings were made both for the original examiners' judgments and the blind evaluators or judges.

Finally, six categories of outcome data from each study were recorded:

1. guilty/deceptive subjects judged correctly;
2. guilty/deceptive subjects judged incorrectly (i.e., judged nondeceptive);
3. guilty/deceptive suspects judged inconclusive;
4. innocent/nondeceptive subjects judged correctly;
5. innocent/nondeceptive subjects judged incorrectly (i.e., deceptive); and
6. innocent/nondeceptive subjects judged inconclusive.

Categories 2 and 5 are the false negative and false positive rates, respectively.

FINDINGS AND DISCUSSION

Three questions are of particular importance to an assessment of polygraph validity useful to policymakers:

1. Are polygraph examinations valid?
2. Given the wide range of outcomes reported across studies, what accounts for their variability?
3. How generalizable are the results of studies to the current and proposed uses for national security purposes?

To answer the first question, data from the available field studies were analyzed to ascertain whether polygraph examination accurately differentiate deceptive suspects from nondeceptive subjects. For this analysis, the outcome frequencies for each category were converted to percentages, and average percentages within each category were calculated. A measure of predictive association (lambda$_b$, see 64,73) was also calculated, although the use of a single measure is very limited due to the wide variability in study design.

The lambda$_b$ index shows the proportional reduction in the probability of error in predicting one category (in this case, deception) when a second category (in this case, polygraph examination results) is known. If the information about the second category does not reduce the probability of error in predicting the first category at all, the index is zero, and one can say that there is no predictive association. On the other hand, if the index is 1.00, no error is made in predicting one category from another, and there is complete predictive association. Essentially, lambda provides an index that translates to the percent improvement over the base rate and indicates the percent improvement in prediction when the polygraph examinations are considered versus no further information. There is almost no direct research on the percent improvement of the polygraph over other forms of investigation (cf. 207). The results of this analysis of predictive association are shown in tables 4 and 5. The average lambda$_b$ across studies is 0.65, which means that, on the average in these field studies, the polygraph diagnosis reduced 65 percent of the error of chance prediction. The lambda$_b$ for individual studies ranged from 0.13 to 0.90.

To summarize, the analysis of the 10 field studies included in the analysis indicates that while polygraph examinations using CQT in criminal investigations detect deceptiveness and nondeceptiveness better than chance, there is also what in some cases might be considered a high error rate, particularly for nondeceptive subjects. The one study which tested the validity of the relevant/irrelevant question technique (the general question test (GQT) portion of the Bersh study) also detected deceptiveness and nondeceptiveness better than chance.

Variation Among Studies

As implied in the introduction to this section, the use of a single statistic or summary number to describe the results of field tests of validity may be misleading. As shown in table 3, although the field studies of polygraph validity are similar in that almost all of them tested control question techniques in criminal investigations, they differ in operationalizations of ground truth and type of examiner decision. The result is that there is a great deal of variability in the results of studies. Correct guilty detections range from 70.6 percent in one condition of the Bersh study to 98.6 percent in a condition of the Wicklander and Hunter study. Correct innocent detections are even more variable, ranging from a low of 12.5 percent in the Barland and Raskin judicial outcome study to a high of 94.1 percent in one condition of the Bersh study. Table 5 also indicates the range of incorrect judgments and inconclusives among studies. False negatives range from 29.4 percent of the Bersh study to zero percent. False positives range from 75 percent in Barland and Raskin (22) to zero percent in two studies. Inconclusives range from zero to 25 percent. This section compares studies that used comparable operationalizations of construct and criterion validity in an attempt to discover reasons for the range of results. However, even using this method results in considerable variability. The main point, however, is that no field studies exist to directly test the situations for which DOD and the President propose to expand polygraph use.

Table 4.—Mean Detection Rates as a Percentage of Total in Field Studies

	"Ground truth"				
Examiners or evaluators' diagnosis	Percent guilty		Percent innocent		
	Mean	S.D.	Mean	S.D.	
Deceptive	49.3	(12.7)	8.2	(7.2)	57.5
Nondeceptive	5.8	(5.1)	32.7	(16.7)	38.5
Inconclusive	2.0	(3.0)	2.1	(2.5)	4.0
	57.1		43.0		100 %

NOTE: lambda$_b$ = 0.65.
S.D. = standard deviation.

Studies Using Panel Criterion and Examiners' Decisions

Both Bersh (29) and Barland and Raskin (22) used a panel to establish the criterion for validity in their studies. The makeup of the panels and the polygraph scoring systems were similar in each study. In the Bersh study, which validated polygraph examinations conducted by military examiners, the panel consisted of four Judge Advocate General (JAG) Attorneys; Barland and Raskin's panel consisted of two criminal defense attorneys, two criminal prosecuting attorneys, and a judge. The examiners in the Bersh study used either GQT (a version of R/I) or the zone of comparison (ZOC) technique; for all but one subject in Barland's study, the Federal ZOC control question technique was used and results evaluated using the Army scoring procedure. Assuming the accuracy of the panel's decisions, the two studies' results are strikingly different. Barland and Raskin attained accuracy rates of 91.5 percent for guilty

Table 5.—Outcomes of Field Studies of Validity

	Guilty				Innocent					
	Number of cases	Correct	Incorrect (false negative)	Inconclusive	Number of cases	Correct	Incorrect (false positive)	Inconclusive	Total number of cases	Lambda
Bersh (29) (panel of 4) GQT unanimous	32	96.9%	3.1%	0[a]	36	88.9%	11.1%	0[a]	68	
ZOC unanimous	38	89.5	10.5	0	51	94.1	5.9	0	89	
Average unanimous	70	93.2	6.8	0	87	91.5	8.5	0	157	0.84
Majority (ZOC and GQT)	34	70.6	29.4	0	25	80.0	20.0	0	59	0.82
		81.9	18.1	0		85.6	14.3	0	216	
Horvath and Reid (84) (1 examiner, 10 examiner/evaluators)	20	85.0	15.0	0[b]	20	90.5	9.5	0[b]	40	0.[illegible]
Hunter and Ash (85)[c]—(1 examiner, 7 examiners/evaluators)	10	87.1	11.4	1.4	10	86.4	14.1	0	20	0.74
Slowick and Buckley, (155)[d]—(random selection; 7 examiner/evaluators)	15	84.0	15.3	0.7	15	90.7	6.6	2.7	30	0.77
Wicklander and Hunter (205)—(2 examiners/6 evaluators)										
PG+[e]	10	98.6	1.3	0	10	86.6	8.3	5.0	20	
PG		90.0	8.3	1.6		86.6	5.0	8.3		
Average		94.4	5.0	1.0		86.6	6.7	6.7		0.88
Horvath (82)[f] (10 examiner/evaluators)										
Verified cases	28	77.1	22.9	0[g]	28	51.1	48.9	0[g]	56	0.28
Davidson (47)[h]—(random selection 7 examiners/evaluators)	10	90.0	10.0	0	11	91.0	0	9.0	21	0.90
Raskin, [d,i]	12				4				16	
(1 examiner, 25 evaluators) Numerical		91.7	0	8.3		75.0	0	25.0		0.75
Nonnumerical		83.3	8.3	8.3		25.0	50.0	25.0		
Barland and Raskin (22) (1 examiner, panel of 5)										
Panel[j]	47	91.5	0	8.5	17	29.4	52.9	17.6	64	0.29
Judicial outcome	33	90.9	0	9.1	8	12.5	75.0	12.5	41	0.13
Kleinmuntz and Szucko (92)—(5 examiners/evaluators)[k]	50	75.0	25.0	0[l]	50	63.0	37.0	0[l]	100[b]	0.38

[a] Data for inconclusives not reported; inconclusives appear to total 27 (243 initial N−216 decisions reported).

[b] Examiner/evaluators were not allowed to judge charts as inconclusive as to overall deceptiveness. In another type of analysis on a question by question basis, judgments of doubtful or inconclusive were allowed.

[c] Average of two blind chart analyses spaced at least 3 months apart, done by same examiners.

[d] Average frequencies divided by number of examiners.

[e] PG+ indicates evaluators had access to written information in addition to polygraph charts (e.g., case details, subject behavior during examinations, etc.). Both PG only and PG+ examinations were done by the same examiners 2 months apart.

[f] Excludes Horvath's analysis of 28 unverified cases; because there is no criterion, this is a reliability study.

[g] There were apparently 15 (1.3 percent) inconclusive judgments out of 1,120 total judgments (10 examiners × 112 cases) which the author excluded from further analyses.

[h] Majority decision only.

[i] Seven examiners used numerical scoring; 18 used nonnumerical scoring procedures.

[j] Excludes 28 cases for which the panel was unable to come to a decision as to guilt or innocence.

[k] Decisions were based on one polygraph chart: standard practice generally employs at least three. Also, the evaluations were made by students with little polygraph experience.

[l] Inconclusives were not allowed.

NOTE: GQT = general question test.
ZOC = zone of comparison.

and 29.4 percent for innocent subjects; comparable figures in Bersh's study are 70.6-percent guilty correct and 80-percent innocent correct. It is not clear why there should be this variation, although differences in the nature of the cases, the completeness of the case files, and sample selection may account for some of the differences.

In the Bersh study, cases were initially drawn at random from a pool of criminal investigations conducted by the three military services over a period of 3 years (1963-66); then, any cases which had been judged "indeterminate" by the original polygraph examiner were eliminated. In addition, after polygraph charts were removed from the investigative files, a preliminary panel of judges eliminated from the sample all files containing insufficient evidence to warrant a positive determination of guilt or innocence. Only those cases which resulted in a unanimous decision by the initial JAG panel were retained in the validation sample. Altogether, one-quarter of the cases (80 cases out of 323) were eliminated because of insufficient evidence. This figure does not include the number initially eliminated on the basis of inconclusive polygraph examinations.

In Barland and Raskin's (22) study, the initial pool of subjects consisted of 102 (nonmilitary) criminal suspects referred to Barland by police, defense or prosecuting attorneys. These cases represented the entire population of Barland's cases at that time. Then, 92 of these 102 cases were retained for further analysis on the basis of independence (a case was considered independent where two or more subjects had not been examined regarding the same crime). In one respect (the fact that there was only one examiner), Barland and Raskin's sample was less variable than Bersh's. However, Barland and Raskin did not eliminate from consideration indeterminate examinations. Neither, and perhaps more importantly, did Barland and Raskin eliminate cases in which investigative files without the polygraph were inadequate. As Barland (17) points out, many of the investigative files that were given to the panel were incomplete. The files had been compiled by inexperienced student assistants who often did not know where to obtain necessary information. The officials responsible for providing the information were, more often than not, unavailable or, when they were available, unable to recall the details of a crime. In many cases, few details were available. As a result, one-third of the 92 cases were judged inconclusive by the panel merely on the basis of the investigative files. The figures reported in table 5 are for 64 of the original 92 cases.

It is not clear why there should be an inverse relationship between accurate detection of guilty and innocent suspects in the two studies. It may be that both the panel and the examiner in the Barland and Raskin study consistently tended to presume guilt in the absence of any a priori base rate (see 28,160). The cases in the Bersh study, on the other hand, were initially selected to be equally distributed among deceptive and nondeceptive cases. It is not reported whether the panel was aware of the base rate in the Bersh study.

Studies Using Confession as a Criterion and Blind Evaluations

The remainder of the field studies analyzed tested the validity of polygraph testing by comparing the blind evaluations of polygraph examiners against a criterion of verification by confession. Two exceptions are Barland and Raskin's judicial outcome analysis and one condition in the Wicklander and Hunter study. The confession studies vary somewhat as to source of verified files. The Horvath and Reid, Hunter and Ash, Slowick and Buckley, Wicklander and Hunter, and Kleinmuntz and Szucko studies all used files from polygraph testing firms. Horvath's cases came from police files, Davidson's from military cases, and Raskin's from the Barland cases reported in Barland and Raskin (22; discussed above). The first four studies used files from the firm of John E. Reid & Associates and involved various criminal offenses. The firm used by Kleinmuntz and Szucko is not identified; all of their cases involved theft.

In the first four studies, blind examiner evaluators also came from John E. Reid & Associates. The Reid studies did vary with respect to case selection. Only one study (Slowick and Buckley) reports random selection of cases; in other studies, the cases of only one or two examiners were used. Horvath's (82) blind evaluators were field-trained examiners with a median of 3 years experience, all of whom specialized in conducting polygraph examinations for police agencies. The 25 evaluators in the Raskin (133) study were volunteers who had trained in a variety of places.

The results of the Reid studies do not vary substantially. The greatest deviation from the mean occurred in one condition of the Wicklander and Hunter study in which examiner/evaluators were given additional information about the suspects (verbal and nonverbal behavioral indicators, demographic information) and the cases. This difference, however, was not statistically significant. Even so, it may be reasonable to consider it separately from the other Reid studies, because of the extra information available to evaluators. In the Reid studies, guilty correct identification rates ranged from 84 to 87.1 percent, with an average of 86.5 percent (excluding the 98.6-percent Wicklander result; 88.9 percent including it). The innocent correct rates in the Reid studies range from 86.4 to 90.7 percent with an average of 89 percent. There is no difference when the Wicklander and Hunter condition is included.

An additional difference of note among the Reid studies concerns the false negative rate, which is highest in the studies which either used random selection of cases (Slowick and Buckley) or eliminated the most clear-cut charts from their original selection (Horvath and Reid). There is no apparent explanation for the variation in false positive rates in the Reid studies, which ranged from 5 to 14.1 percent.

The Davidson study results are basically similar to those of the Reid studies, except for the absence of false positives. However, the study should be interpreted with caution as one-third of the originally (randomly) selected sample was not able to be used.

The Horvath (82) and Kleinmuntz and Szucko (92) studies have the lowest accuracy rates. As with the Barland and Raskin (22) study, the low accuracy rate may be related to the fact that Horvath selected his sample from police files. Perhaps, police records of verification are not reliable, or have greater variability than those of polygraph firms.

Barland (17) has suggested a number of reasons why Horvath's results are lower than the Reid studies. One reason is that the blind reviewers did not have access to "special charts" administered in 32 percent of the cases, primarily to subjects the original examiner considered deceptive; these charts were removed from the files before being reviewed by blind examiners. According to Barland, Horvath's original examiners had been 100 percent correct in their judgments. A second reason is that, as noted above, police examiners were used instead of private examiners; the difference between the two kinds of examiners is not explained further. Yet a third reason, which Barland (17) believes may be the most important in terms of false positives, is that a number of victims and witnesses were included in the sample (i.e., were subjects). According to Barland (17), one theory of detection of deception predicts that innocent victims or witnesses may react emotionally during a polygraph examination because they experienced or witnessed the event regardless of whether they are telling the truth about specific details of the incident. An analysis of the Horvath data suggested by Barland, comparing results for victims and witnesses with those for suspects, would be of interest (see Giesen and Rollison (61) for a comparison of innocent associations with guilty knowledge).

Despite the generally anomalous results of Horvath's (82) study, an interesting finding may help to account for the results of the Kleinmuntz and Szucko (92) study. Horvath found that suspects in crimes against property were less detectable than suspects in crimes against persons. This may be because crimes against persons are likely to have a greater amount of affect associated with them, and are, thus, more physiologically detectable. Barland and Raskin (22), on the other hand, found no differences by type of crime. As noted previously (see table 3), Kleinmuntz and Szucko's (92) study selected only cases from the files of a polygraph firm involving crimes of theft. However, although the crimes against property hypothesis is suggestive, it may not fully explain the difference between Kleinmuntz and Szucko's and similar studies. The Davidson study, for example, only used theft cases, and it has a "0" false positive rate (although it has a substantial inconclusive rate). Analyses of other studies by crime type would be informative, although the number of cases would probably be too small for a meaningful analysis.

Szucko (159) has suggested that one possible reason his results are so different from other polygraph firm studies' results, is that the individual who selected the charts in the Kleinmuntz and Szucko study could not read polygraph charts. Therefore, case selection may have been more variable than in some of the other studies. Alternative explanations are that: 1) Kleinmuntz and Szucko only evaluated one chart for each subject (at least three is standard); and 2) their evaluators were examiner-trainees at the end of their internship period, not experienced examiners* (see 91).

Studies Using Judicial Outcomes and Original Examiners' Results

Bariand and Raskin's (22) analysis using judicial outcomes as a criterion has the lowest accuracy rate for innocent suspects—a 12.5-percent innocent correct and 75-percent false positive rate. The problems with using judicial outcomes as a criterion have already been referred to, in particular, the fact that the judicial outcome is not a highly accurate measure of guilt because of such characteristics of the legal system as the necessity for proof beyond a reasonable doubt, and the prevalence of plea bargaining. These problems are illustrated here by the fact that only 41 of Barland and Raskin's original 92 cases were resolved by the criminal justice system. Again, there is clearly greater agreement on guilty subjects.

---

*Some maintain that the evaluators in Kleinmuntz and Szucko's study were even less experienced than that.

OTHER CONSIDERATIONS

Although the analysis above demonstrates that polygraph testing is better than chance at differentiating deceptive from nondeceptive subjects in criminal investigations, what might be considered as substantial false positive and false negative rates are obtained in several investigations. Although it is not possible to determine a "scientifically" acceptable rate of correct or incorrect judgments, clearly if error rates are between 10 and 25 percent, a large number of incorrect decisions would be made if the polygraph were widely employed. The base rate of guilt in actual situations may further complicate matters. It is not clear from the field studies conducted so far how many suspects were involved in the cases selected for polygraph testing, but if there were a large number of suspects, more false positives could be expected (see ch. 7).

Also problematic is the wide variability in accuracy rates across studies. Although some differences can be explained methodologically, other differences cannot. Of perhaps even greater importance than the accuracy rate variability and error rate problems is the observation that field studies of polygraph testing have only been conducted in criminal investigations. As is discussed more fully in chapter 6, criminal investigations may generate different levels of affect. In addition, different kinds of subject groups may be the focus of expanded Government use of polygraph testing. Only two field studies can be identified that relate directly to polygraph testing in the national security area: one by the Director of Central Intelligence (DCI,165) and a second by Edel and Jacoby (51). Neither of these is a validity study but because they are the only field studies with any relevance to national security, they will be described below in some detail. An analog study of counterintelligence screening (16) is discussed in chapter 5.

The DCI study consisted of a survey of 12 Government agencies (not including the National Security Agency (NSA)). The study was conducted to evaluate the relative effectiveness of various means of conducting background investigations for purposes of applicant screening and security clearances for current employees. Background investigations are conducted through the use of personnel interviews, interviews with present and former neighbors, checks of educational and work records, and checks with a consortium of other national agencies (the so-called National Agency Check). Of the agencies surveyed, only the Central Intelligence Agency (CIA) used the polygraph to conduct background investigations.

In the 4-month period covered by the study, CIA conducted 507 background investigations. Of these, adverse information arose concerning 47 percent of applicants or other individuals being investigated for security clearances. Thirty-five (83 percent) of the adverse cases were resolved against the individual (i.e., the applicant was not hired or clearance was not granted). In two-thirds of the instances of adverse information resolved against the individual with the use of the polygraph, subjects admitted to the adverse information. The kinds of issues admitted by subjects had primarily to do with drug and alcohol use (e.g., marijuana use, alcohol abuse, abuse of other drugs; approximately 55 percent of the cases) and immoral conduct (e.g., sexual deviance; 24 percent of cases). Four cases involved irresponsibility, a subcategory of which is violation of security regulations, and none involved the loyalty category. It is not clear whether any of the four irresponsibility cases involved violations of security regulations. Three of the eighty-four resolved against cases involved admissions of foreign connections, meaning in this case either that:

1. the subject was not a U.S. citizen;
2. the subject's spouse was not a citizen;
3. relatives were potential "hostages;"
4. alien relatives, "hostage" unlikely; or
5. life abroad cannot be verified.

The seriousness of the wrongdoings was not clear.

The crux of the DCI analysis was the construction of a productivity index for investigative techniques from the CIA data and data from other agencies. Based on the fact that a large number of cases were resolved against individuals by admission, and the polygraph was the "unique source" (165) in all the CIA cases resolved against the subject, DCI tentatively concluded that the polygraph was the most productive of all background investigation techniques. For admissions, for example, the polygraph had an index of 6.59 compared to 0.79 for "administrative screening," 1.08 for "investigative interviews," and 0.28 for "papers only."

Several aspects of the study should be noted. One is that the criteria for case selection and adverse information are not stated. Another issue, noted by the DCI study authors, is that even though the polygraph is reported as the sole source in resolving adverse information, it was only used after a thorough investigation using other sources had taken place. For this reason, it is difficult to assess its effectiveness separately from the effect of a thorough investigation. Furthermore, as a result of being conducted at the end of a background investigation, in this case the polygraph examinations could be considered a confrontation technique rather than an investigative tool, according to DCI. Agencies surveyed by DCI were asked not to include confrontation techniques in their responses. A third problem is that there was no independent verification of the cases that were resolved. Perhaps most important, the effectiveness of polygraph examination cases involving most, if not all (i.e., irresponsibility) of the kinds of adverse information uncovered among applicants in the sample probably cannot be generalized to investigations of unauthorized disclosures.

Edel and Jacoby (51), in a study reported in a leading psychology journal, tested the reliability of polygrapher judgments of physiological responsivity in applicants for positions with "a large Government agency." Forty cases were randomly selected from the agency's applicants in 1966. Ten practicing polygraph examiners acted as actual examiners in four cases each and raters in eight additional cases. In each case, examiners (raters) judged three physiological responses to each interview question as either "no specific reaction" or "a specific physiological reaction." The rate of agreement between examiners and raters as to whether a physiological reaction took place averaged 96 percent.

Of course, as the authors note, demonstrating consistency among examiners "is not equivalent to demonstrating consistency in interpretations based on these physiological reactions" (51). For example, responses were not differentiated for relevant v. irrelevant questions. Therefore, although Edel and Jacob's study indicates that the examiners in the Government agency can reliably detect physiological reactions, whether these physiological reactions indicate deception among applicants for positions in Government agencies has not been tested. Because of the potential adverse consequences for employment applicants (particularly in Government agencies where there is interagency checking (see, e.g., 165)), such tests have substantial practical significance.

CONCLUSIONS

Although there is some evidence from available field studies that polygraph testing is effective in detecting deception by guilty criminal suspects, there is also what in some cases might be regarded as a substantial error rate. This is particularly so for innocent subjects. There appears, as yet, to be no scientific field evidence that polygraph examinations can be effectively used to investigate unauthorized disclosures or that they represent a valid test to prescreen or periodically screen Government employees. Results of field studies are subject to additional problems of interpretation because of inadequate measures of ground truth.

The following chapter reports on the effectiveness of polygraph testing demonstrated by analog studies. As will be shown, the construct and criterion components of validity are stronger in analog studies, but because of problems with external validity, they do not provide evidence about actual polygraph testing that is as direct as that from field studies. Nevertheless, reviewing such evidence is necessary to assess both the present and potential use of polygraph testing.

Chapter 5

Review and Analysis of Polygraph Analog Studies

Chapter 5

Review and Analysis of Polygraph Analog Studies

INTRODUCTION

Analog studies, for purposes of the present analysis, are investigations in which field methods of polygraph examinations are used in simulated criminal or other situations. Such studies investigate either "mock" crimes set up by an experimenter (with the knowledge and collaboration of subjects) or actual small crimes "induced" by the experimenter. Such analog studies are not actual criminal investigations and subjects are usually aware that they are participants in polygraph research. Analog studies differ from other laboratory studies of polygraph testing in that they simulate actual field examinations. However, in analog studies, typical components of field examinations are replicated to the extent it is possible to do so. Such studies test the validity of various polygraph techniques under controlled conditions. In chapter 4, the results of a systematic review of field studies of validity were presented. In the present chapter, a similar analysis of analog studies is presented. As with the field studies, the studies concern the use of polygraph examinations for investigation of crimes. The two exceptions (16,43) use analogs to the type of relevant/irrelevant (R/I) question technique typically used in the personnel screening situation.

The present chapter is organized as follows: first, the characteristics of analog studies and the varieties of ways in which they differ from field studies are discussed. Then, the criteria used for including studies in the analysis are described. The coding procedure, which is essentially the same as that used to code the field studies, is described briefly. Analog studies of the control question technique (CQT), guilty knowledge technique (GKT), and personnel screening examination are then reviewed. The findings of a statistical analysis of the analog studies complete the chapter.

CHARACTERISTICS OF ANALOG STUDIES

The "crimes" utilized in analog studies in order to establish ground truth have taken different forms. For the most part, they are "mock crimes;" i.e., crimes in which subjects know they are "role playing" at being criminals for purposes of an experiment. Mock crime studies may be further differentiated by whether or not the experimenter controls the guilt or innocence of research participants. In some studies, subjects know that the crime is part of the experimental situation but they are more or less free to go through with the crime or not. Two analog studies have utilized actual small crimes. In these studies, apparently real situations were embedded in an experimental situation in which subjects were given an opportunity to commit a crime or not.

The consequences of failing a polygraph examination (e.g., a possible prison sentence) cannot be replicated in the laboratory. In analog studies, punishment takes such forms as losing the chance for a monetary reward. Some researchers have experimented with other punishments such as electric shock (105) or the threat of shocks (35). The analog studies that use real crimes provide another alternative, in that subjects can be threatened with real punishment (e.g., academic sanctions for cheating on an examination). In still other cases, subjects are led to believe that "stable" individuals can avoid detection.

Analog studies represent, thus, a "tradeoff" to the investigator interested in polygraph testing validity. On the one hand, because the researcher sets up the crime, ground truth is known; and because "ground truth" is established, analog studies are superior to field studies in terms of criterion validity. Furthermore, they provide the investigator with more control of the polygraph situation and conditions of testing. The experimenter can select particular subject groups, can standardize testing procedures for all subjects, and can systematically vary guilt or innocence. With this control, the experimenter can also directly compare the effects of variations in polygraph techniques, physiological measures, information given to subjects, and scoring methods.

On the other hand, although analog studies have greater criterion validity and offer greater experimental control, their use as indicators of polygraph testing validity is potentially problematic. The reasons have to do primarily with external validity (20,136; see, also, 1,7,108); i.e., the crime situation differs, the testing situations in the field and the laboratory differ, the training of the examiners differs, the subject population differs, and, apparently most important, the consequences for "suspects" differ dramatically between the field and the laboratory. In addition, in analog studies, the questions and question techniques most often are not tailored to individual subjects. In actual criminal field investigations, case information about the crime and the subject usually provides a basis for tailoring questions.

Numerous specific differences can be noted. Perhaps most importantly, the laboratory crime and the consequences of detection are much less serious. In addition, in an analog study, demand characteristics (which suggest to the subject desirable responses) may create a somewhat different polygraph situation than found in typical field situations (20). In terms of factors that may increase validity of analog studies, there is some evidence that laboratory researchers are, in general, able to use more sophisticated and stable equipment than portable machines often used in the field (136). On the other hand, examinations in analog studies are often conducted by researchers who are primarily psychophysiologists (e.g., 49) or psychologists (43) with only limited training in field techniques. Field examinations, in contrast, are conducted by individuals whose primary training is as polygraph examiners and who are usually experienced. This would suggest that field examinations may be more accurate.

The characteristics of subjects who participate in analog studies also vary from subjects in field studies. Several use college students, others recruit community members through the newspaper, one uses police candidates, and another prison inmates. In many studies, subjects are probably better educated and more highly socialized than the average field examinee. In the case of student subjects, they are probably younger on the average and from a higher social class as well. Raskin (132) notes that analog studies using students yield a lower accuracy rate than other studies. As will be discussed below, this may be due to subject differences between field and analog studies because a realistic fear of failure does not play a central role for subjects. The consequences of failure for analog studies are usually minimal in contrast to typical field investigations.

Study Selection

For present purposes, studies were only included as analog for the primary analyses if they employed actual field polygraph techniques to detect deception or concealed information, and if the studies pertained to some use of polygraph testing in the real world. The studies selected are listed in tables 6 and 7. Studies of components of the polygraph examinations, such as studies which used only card tests (97,101), number tests (120), or tests concerning concealed personal information (e.g., parents' first name; see, e.g., 106) were not included.

In addition, studies were excluded because their primary focus was on a theoretical factor thought to affect validity, such as variability in physiological recordings (45), nonstandard means of interpreting such recordings (163), or the role of "lying" (96). Such studies will be referred to as laboratory investigations and are distinguished from analog studies.

Analog studies of the guilty knowledge test (GKT) have been included, although analyzed separately, because this form of the polygraph examination represents an alternative proposed for

Table 6.—Outcomes of Control Question Analog Studies of Validity

	Guilty				Innocent					
	Number of subjects	Correct	Incorrect (False negative)	Inconclusive	Number of subjects	Correct	Incorrect (False negative)	Inconclusive	Total number of subjects	Lambda[b]
Barland and Raskin (21)	36	63.9%	8.3%	27.8%	36	41.7%	16.7%	41.7%	72	0.47
Podlesny and Raskin (127)	20	69.0	16.0	15.0	20	91.0	4.0	5.0	40	0.75
Raskin and Hare (137)	24	87.5	0	12.5	24	75.0	4.0	20.8	48	0.83
Rovner, et al.[a] (143)	36	77.8	8.3	13.9	36	80.5	13.9	5.5	72	0.72
Kircher (89a)	50	60.0	4.0	36.0	50	76.0	2.0	22.0	100	0.72
Dawson (49)	12	91.7	0	8.3	12	58.3	25.0	16.7	24	0.33
Widacki and Horvath[b,i] (207)	20	90.0	5.0	5.0	—	—	—	—	20	—[k]
Bradley and Janisse (35)	96	—	—	—	96	—	—	—	192	0.33
EDR		60.4	13.5	26.0	—	58.3	9.4	32.3		
Heart rate		35.4	20.8	43.8	—	33.3	19.8	46.9		
Szucko and Kleinmuntz[c] (160)	15	71.3	28.7	d	15	49.3	50.7	d	30	0.22
Ginton, et al. (62)	2	100.0	0	0	13	84.6	15.4	0	15	0.00[j]
Honts and Hodes (75)[e]:										
No countermeasures	9	67.0	0.0	33.0	12	33.0	17.0	50.0		0.25
Countermeasures	19[f]	58.0	5.5	36.6	—[g]	—	—	—		—
	28				12				40	
Honts and Hodes (76):										
No countermeasures	19	84.2	0.0	15.8	19	31.6	15.8	52.6		0.53
Countermeasures	19	36.8	26.3	36.8	—[g]	—	—	—		—
	38				19				57	
Heckel, et al. (74)[i]:										
Normals	—[h]				5	100.0	0.0	0.0		—[k]
Nondelusional psychiatric	—[h]				5	70.0	10.0	20.0		
Delusional psychiatric	—[h]				5	45.0	35.0	20.0	15	
Hammond (74a)	32	71.9	3.0	25.0	30	40.0	20.0	40.0	62	

[a] Summed across conditions.
[b] Examiner's task was to detect the one guilty person in each of 20 groups of four suspects.
[c] Based on ratings of 5's on a 1 to 8 scale of certainty of nondeception/deception.
[d] Examiners were not allowed to categorize an examination as inconclusive.
[e] Original subject assignments, 12 to each of 4 (including 2 countermeasure) conditions. A total of five countermeasure subjects were eliminated from the analysis of results for guilty subjects for failure to follow countermeasure instructions. Three no countermeasure subjects were eliminated for spontaneously using countermeasures.
[f] Nine guilty subjects used pain countermeasures (tongue biting) and 10 used a muscle (toe pressing) countermeasure.
[g] Innocent subjects used no countermeasures.
[h] There was no guilty condition.
[i] Not included in analysis reported in table 8.
[j] Lambda is a poor statistic when the base rate is skewed.
[k] Lamba was not calculated when only guilty or innocent subjects were used.

Table 7.—Outcomes of Guilty Knowledge Analog Studies of Validity

	Guilty				Innocent					
	Number of subjects	Correct	Incorrect (false negative)	Inconclusive	Number of subjects	Correct	Incorrect (false positive)	Inconclusive	Total number of subjects	Lambda$_b$
Lykken[a] (105)	50	88.0%	12.0%	—	48	100.0%	0%	—	98	—[c]
Davidson (46)	12	91.7	8.3	—	36	100.0	0	—	48	0.92
Podlesny and Raskin (127)	10	80.0	20.0	0	10	80.0	0	20.0	20	0.80
Balloun and Holmes (12)	18			—					34	
Test 1		61.1	38.9	—		87.5	12.5	—		0.44
Test 2		16.7	83.3	—		93.7	6.3	—		
Giesen and Rollison (61)	20	95.0	5.0	—	20	100.0	0	—	40	0.95
Bradley and Janisse (35)	96			—	96				192	
EDR		59.4	40.6	—		88.5	11.5	—		0.38
Heart rate		44.8	55.2	—		82.3	17.7	—		
Timm (163)	237								237	
Liberal cutoff		80.8	19.2	—	—[b]	—	—	—		—[d]
Conservative cutoff		70.4	29.6	—	—	—	—	—		—

[a] Frequencies for detection of two mock crimes were combined.
[b] There were no innocent subjects.
[c] Lambda cannot be calculated because crimes were not reported separately.
[d] Lambda cannot be calculated with only one condition.

use in the field (92,107,108), even though it has not been put into general practice.

Description of Studies

The following sections discuss each of the analog studies organized into three categories according to questioning technique. The discussion of CQT analog studies is first. Studies of CQT represent available studies, much like the case for field investigations (see ch. 4). Six studies of the concealed information or GKT and two of R/I follow. In only one study (16), involving the R/I technique, were subjects Government employees. The results of individual studies are summarized in tables 6 (CQT) and 7 (GKT). The description of the studies is followed by a systematic statistical analysis of the results of the CQT and GKT studies. The R/I studies were not analyzed as a group because of the paucity of studies.

Essentially, as shown in tables 6 to 9 the analysis of the analog studies yields conclusions similar to those of the field study analysis—i.e., although there is a greater-than-chance probability of detecting deceptive and nondeceptive subjects, there is what might be regarded as a significant error rate, and a great deal of variation across studies. However, as has been found in some reviews (1,7), analog studies of CQT had lower accuracy rates than field studies of CQT.

In the studies detailed below, some experiments also tested the effect of factors hypothesized to have an effect on validity. For example, Barland and Raskin (22) examined the effect on validity of different types of feedback about the polygraph, and Dawson (49) investigated the effects of countermeasures. These factors are examined more systematically in chapter 6; the emphasis of the present chapter is on the validity of different forms of polygraph examinations.

Table 8.—Mean Detection Rates as a Percentage of Total in Analog Studies of Control Question Technique

	Ground truth	
	Percent guilty	Percent innocent
Examiners' diagnosis	Mean	Mean
Deceptive	33.0	6.8
Nondeceptive	5.4	27.9
Inconclusive	13.4	13.5
	51.8	48.2

NOTE: lambda_b = 0.43.

Table 9.—Mean Detection Rates as a Percentage of Total in Analog Studies of Guilty Knowledge Test

	Ground truth	
	Percent guilty	Percent innocent
Examiners' diagnosis	Mean	Mean
Guilty	27.9	2.2
Not guilty	17.3	52.6
Inconclusive	0	0
	45.2	54.8

NOTE: lambda_b = 0.70.

CONTROL QUESTION TECHNIQUE

Fourteen analog studies of the control question technique were located. The largest group of these studies emanate from the research program of Professor David C. Raskin at the University of Utah. The remaining studies were conducted at a number of settings in the United States and elsewhere. Raskin and colleagues have conducted a systematic analog research program, and these studies are described as a group. Other researchers have published individual studies testing specific hypotheses relevant to the validity of the polygraph. A description of these studies follows discussion of the University of Utah studies.

University of Utah Studies

Despite longstanding controversy about polygraph validity, the first research project conducting an analog study that simulated field polygraph techniques was not conducted until the 1970's (136). It was then that an ongoing research program headed by Professor Raskin at the University of Utah began to study the validity of the polygraph through analog experiments. In addition, these studies also examined the relationship to validity of different polygraph techniques (e.g., the stimulation test), different physiological measures, different methods of assessing the results, different types of information provided to subjects, and different subject and situation factors that could potentially affect polygraph validity.

The experiments conducted by Raskin and colleagues use similar procedures to set up the mock crime and to conduct polygraph testing. In each of their studies, subjects are randomly assigned to an "innocent" condition or to a "guilty" condition. The mock crime is the theft of a small amount of money or a valuable object from a desk in a nearby room. To increase their motivation, subjects are offered a financial bonus for convincing the examiner they are innocent. In the testing the examiner employs the Federal zone of comparison (ZOC) control question technique, including a pretest interview. A numerical field scoring method developed by the Utah group (21) is used to make the diagnosis of truthfulness or deception.

Barland and Raskin

In the initial analog study using CQT (21), 72 student "guilty" and "innocent" volunteers were randomly assigned to one of three "feedback" conditions. The positive feedback subjects were instructed that the polygraph was effective, the negative feedback students were told that the machine was not working properly, and the other students received no feedback. Subjects then underwent a complete polygraph examination including a pretest interview. The Federal version of the ZOC technique was employed, with standard control questions used for all subjects. On average, the CQT identified 53 percent of all subjects correctly. Twelve percent were identified incorrectly and 35 percent of the examinations were inconclusive. Of the errors, three (4 percent of the entire sample) were false negatives and six (8 percent) were false positives.

Podlesny and Raskin

Podlesny and Raskin (127) conducted a more extensive experiment to examine the accuracy of CQT using three different types of control questions. They also tested the accuracy of behavioral observations of the examinee (80,139) in detecting deception, since this type of information is used in many field examinations and could possibly affect the validity of the technique (107,108). They compared as well the capability of different physiological measures in differentiating between guilty and innocent subjects. A GKT was also conducted with 20 subjects (see below).

In Podlesny and Raskin's study, subjects were community members who responded to newspaper advertisements. The experimenters drew from the Reid method in their design of the pretest interview (see ch. 2). One experimenter asked the subjects three questions from Reid and Horvath's structured pretest interview designed to provoke the subjects into displaying "behavioral symptoms" of deception (80,139).

During the polygraph examination they included two special types of control questions among the set of questions asked of the subjects. One was a "guilt complex question," which asked the subject if he committed a fictitious crime of the same nature as the real crime. In this study, the guilt complex question was, "Did you take that watch from room 702?" (127). There was, of course, no watch stolen from room 702. The experimenters also varied the wording on some of the control questions, so that half the subjects received "nonexclusive" and half "exclusive" control questions.

In the pretest interview, the examiners followed the usual field procedure of reviewing the control questions with the subjects, and the questions were adjusted until they elicited a "no" response. The control question polygraph test then took place, with three or more charts obtained from each subject, although only the first three were used in the objective scoring. Immediately after testing, the first three charts obtained were scored blind on electrodermal response (EDR), respiration, and cardio measures. Later, an independent rater scored the tests, using the numerical scoring system devised by Barland and Raskin (21). The experimenters also used objective measurements of all physiological response measures with the aid of computers and persons who had no knowledge of the field evaluations or treatments administered. The experimenters used the decisions made by the independent blind evaluator to assess the validity of the polygraph test. This was, however, equivalent to using the polygraph examiner's decision, because the independent rater and the examiner agreed on 100 percent of their decisions.

The results for both types of control questions combined (with an inconclusive zone used) were 80 percent correct, 10 percent incorrect, and 10 percent inconclusive. There were three false negatives (8 percent) and one false positive (2 percent). The accuracy of CQT differed depending on whether exclusive or inclusive control questions were used. When the exclusive control questions were used, 85 percent of the subjects were identified correctly, 5 percent incorrectly, and 10 percent inconclusively. Of the assessments of the 20 subjects in this condition, one (5 percent) was a false negative and there were no false positives. When nonexclusive control questions were used, 75 percent were correct, 15 percent incorrect, and 10 percent inconclusive. Of these 20, two (10 percent) were false negatives and one (5 percent) was a false positive. The evaluative scores for each physiological component were analyzed to determine if the scores differed between guilty and innocent subjects. Only the EDR and plethysmograph scores yielded significant differences.

Behavioral observations, by themselves, yielded a significant number of correct decisions, but this differed greatly between innocent and guilty subjects. Of the guilty subjects. 86 percent of the decisions made were correct (25 deceptive, 4 nondeceptive, and 1 inconclusive); however, only 48 percent of the innocent subjects were correctly identified (12 deceptive, 11 nondeceptive, 7 inconclusive). An objective quantitative analysis for each physiological measure was employed to determine if each was effective in discriminating between guilty and innocent. Most of the measures yielded significant discriminations, with the exception of a few of the cardiovascular measures.

Raskin and Hare

A special population, prisoners, especially relevant to the field use of the polygraph, was the focus of a study by Raskin and Hare (137). In their sample of 48 inmates of a Canadian prison, half were selected for high levels of psychopathy, and half for low levels. One purpose of their study was to investigate whether deceptive psychopaths could more easily escape detection than normal subjects (see ch. 6). Overall, assessments of deception from the field evaluations from all charts were 88 percent correct, 4 percent wrong, and 8 percent inconclusive. There were only two errors, both false positives. No significant differences were found between psychopaths and nonpsychopaths, suggesting that a CQT polygraph examination is equally valid for both. Also, a quantitative analysis showed that all the physiological measures were significantly different between guilty and innocent subjects. Psychopathy did not obscure these differences and in some cases enhanced them.

Rovner, Raskin, and Kircher

Rovner, Raskin, and Kircher (143) studied the effect of information and practice on the accuracy of polygraph examinations. Seventy-two subjects recruited from the community took part in this mock crime experiment. One third of the subjects (12 innocent and 12 guilty) were given in-depth information about the polygraph and about countermeasures used to appear innocent (information condition). Another third received this information and underwent two practice polygraph examinations about which they received "feedback" (information and practice condition). The other third had no such intervention (standard). A blind field evaluation performed some time later produced the scores for decisions of guilt or innocence, and for analysis of the physiological responses. Accuracy for the standard group and the information group was identical: 88 percent correct, 4 percent incorrect, and 8 percent inconclusive. But accuracy for the information and practice condition was lower: 62.5 percent correct, 25 percent incorrect, and 12.5 percent inconclusive. There was one error in the standard group and one in the information group—both false positives. The six errors in the information and practice conditions were three false positives and three false negatives.

Kircher

Some of the latest work of the Utah laboratory explores the use of computers in the analysis of polygraph recordings. Kircher (91a) compared the accuracy of a computer decisionmaking process to the accuracy of assessments of a field examiner. The computerized analysis cannot be included in the statistical analysis of this technical memorandum, because it is not presently a field scoring method, but the decisions of an independent evaluator who was used can be. This mock crime study followed the basic procedures of Podlesny and Raskin (127) with 100 subjects from the community. The accuracy of the original examiner was not reported though the results of an independent evaluator were. The independent evaluator, who numerically scored the charts blindly, correctly diagnosed 87 percent of the subjects; misdiagnosed 6 percent; and made a judgment of inconclusive on 7 percent. The six errors were evenly divided between three false negatives and three false positives. In comparison, different computer decision models, on the average, correctly identified 84.9 percent of subjects, misidentified 7.85 percent, and placed 7.2 percent in an inconclusive category.

Other Studies

A range of other studies has been conducted in recent years to evaluate aspects of polygraph test validity. Such studies usually manipulate one or two variables that are hypothesized to be important determinants of polygraph validity. For the most part, these experiments use procedures that are similar to Raskin's mock crime paradigm. Some of the discussion of the procedures in each study is omitted, because they closely follow this paradigm.

Dawson

Dawson (49), for example, focused on the effect of "cognitive countermeasures" on validity. His study was unique in that the subjects were actors trained in the Stanislavsky method of acting, which teaches actors to use their own experience to create emotional states appropriate for a role. Studying the attempts of "method" actors to foil the polygraph may help determine whether guilty subjects can be trained to use cognitive countermeasures to appear innocent (see ch. 6). Dawson was also interested in analyzing separately responses during two distinct phases of the questioning: while subjects listened to questions and while they responded.

Dawson's sample consisted of 24 student actors, half of whom were randomly assigned to the "guilty" group and half to the "innocent" group. They were instructed to use the Stanislavsky method to appear innocent on the polygraph examination. After the mock crime, four charts were obtained from ZOC control question test about the crime. On two of the charts, the subjects were instructed not to respond until they received a signal 8 seconds after a question. This served to separate responding associated with the questions from responding associated with answering. Numerical scoring based on Barland and Raskin's (21) system was done separately on three different types of physiological responses:

1. responses when the answers were immediate;
2. responses during the questions when the answers were delayed; and
3. responses during the answers when the answers were delayed.

Dawson found that the subjects' immediate physiological responses to the questions, whether they were answering immediately or not, led to decisions which were 88 percent correct, 8 percent incorrect, and 4 percent inconclusive (frequencies across two conditions were summed). The delayed answer response yielded a rate of 29 percent correct, 8 percent incorrect, and 62 percent inconclusive. The incorrect decisions made were entirely false positives. A quantitative analysis revealed that the EDR and cardiovascular measures differentiated significantly between innocent and guilty, but respiration did not. The major outcomes of this study suggested that the polygraph was not susceptible to cognitive countermeasures of the sort used by the actors and that scorable responses generally occur immediately after questions.

This experiment does not, however, test cognitive countermeasures in a situation in which the subjects know the essentials of CQT and apply cognitive countermeasures differentially to relevant and control questions. The average criminal subject is likely to attempt cognitive measures naively, but a sophisticated subject—perhaps the type more likely to appear in a national security investigation—may learn cognitive countermeasures along with the knowledge of the control question or other technique.

Widacki and Horvath

Widacki and Horvath (207) designed an experiment to examine the polygraph's efficacy in comparison to other techniques in the mock investigation of a mock crime. They recruited 80 Polish student volunteers and had all of them provide writing specimens, photographs of themselves, and fingerprints. Subjects were then assigned to 20 groups of four subjects each. Within each group, one subject was randomly assigned to be the perpetrator, and the other three were innocent suspects. Each group was thus an "investigative case." Because of this feature of the design, the decisions of guilty and innocent were not independent. Therefore, Widacki and Horvath's findings could not be included in the statistical analysis of the control question analogs and must be considered separately. A similar situation holds for Kubis' (93) mock crime experiment (see below).

The mock crime proceeded as follows: the guilty subject picked up a parcel from one of two persons acting as a "doorkeeper" of a building in the area. The perpetrator gave some experiment-related papers to the doorkeeper and then signed for the parcel. Thus, an eyewitness account (by the doorkeeper), fingerprints, and handwriting specimens were all available. Blind polygraph examinations then were conducted using the Reid control question method (including the examiners' behavioral observations of the subject). Analysis of the three other sources of evidence was carried out.

Widacki and Horvath found that the polygraph produced the most correct decisions (n=18), the fewest (along with handwriting) incorrect decisions (n=1), and the fewest inconclusive decisions (n=1). Widacki and Horvath note, however, that a direct comparison of these four investigative methods may be invalid because the experimental procedures could not ensure a comparable level of quality of evidence for each method (e.g., fingerprints were not detectable in the majority of cases).

Because of its experimental design that had the examiner make decisions on four suspects as a group, the study produces data about the accuracy of the polygraph that is difficult to interpret. But it does shed light on the efficacy of the polygraph relative to other investigatory techniques that might be the alternative. Certainly, it is crucial in policymaking to judge the validity of the polygraph relative to other techniques that would be used in its stead. More research is needed in which the polygraph is compared to other investigatory techniques, and the quality of information across techniques is held constant. Such a comparative analog study would be especially valuable if it included different techniques used in investigations of Federal personnel, such as those reported in the Director of Central Intelligence (DCI) survey mentioned in chapter 4 ("administrative screening," "investigative interviews," etc.).

Bradley and Janisse

Bradley and Janisse (35) studied the effects of two other variables hypothesized to influence the validity of the polygraph: the degree of threat involved in the punishment for being judged guilty, and successful demonstration to the subjects of the technique's accuracy. A mock crime was carried out using procedures similar to those used by Barland and Raskin (21). Subjects were also given a series of stimulation tests. Results of these tests were manipulated such that they made the polygraph test appear perfectly effective, partially effective, and ineffective. In addition, half the subjects were told they would receive a painful electric shock if found guilty, though no shock was ever given.

The degree of manipulated effectiveness had no direct effect on scores, but did tend to increase the accuracy of detection. Threat of punishment did not affect accuracy of detection, although it did have an overall effect on heart rate. EDR and heart rate change were significantly accurate in differentiating guilty and innocent, although another measure, pupil size change, was not.

Honts and Hodes

Two recent analog studies of the Backster ZOC method of testing (76,77) were conducted primarily for the purpose of testing whether polygraph examiners could detect the use of physical countermeasures by subjects. In the first study, subjects were college students who received extra credit toward their final grades for their participation. "Guilty" suspects participated in a mock crime (theft of an examination); innocent suspects were only told of the theft. All subjects were motivated to produce truthful outcomes on the polygraph test by an offer of twice the number of credits if the examiner reported them as truthful.

In addition to participation in the mock crime, 24 of the guilty subjects participated in 15-minute training sessions in which they were told about the theory of CQT and shown how to use either tongue biting (12 subjects) or toe pressing (12 subjects) as countermeasures during presentation of the control questions. They were also instructed to try to relax as much as possible during presentation of the relevant questions.

The actual polygraph examinations took place a week after the theft and training sessions. All guilty subjects were instructed to have the "stolen" examinations with them, presumably to enhance subject involvement. Four charts were obtained from each subject using a standard Backster examination administered by an experienced polygraph examiner. The examiner was aware of the details of the experiment, including a knowledge of the base rates of guilt and the countermeasures that would be attempted, but was blind to the group assignment of individual subjects. At the end of each examination, the examiner made a yes/no decision regarding the subject's use of countermeasures. After all subjects had been tested, the original examiner made a decision as to deception by blindly evaluating the charts using the Backster numerical scoring technique, and made another decision about the use of countermeasures based on inspection of the charts. Charts were also examined and scored by a second examiner who was blind to all aspects of the experiment.

As shown in table 6, while there was a low rate of false negatives (5.5 percent), examiners were not able to make a decision on one-third of countermeasure and no countermeasure guilty subjects, and half of the innocent subjects. There was a 7 percent false positive rate. Examiners were not able to detect the use of countermeasures.

In their second experiment on countermeasures, Honts and Hodes used approximately the same procedures and subject pool, with the exception that subjects were asked to employ both countermeasures simultaneously, were given 30 minutes of training, including a practice session, and were asked to practice at home. A cardio cuff was added to the polygraph instrument, and a card test was conducted prior to the administration of the first test.

Overall, results of the second study replicated the first. The categorizations of the original examiner were 51 percent correct, 14 percent incorrect, and 35 percent inconclusive. Twenty-six percent of the countermeasure subjects compared to none of the guilty/no countermeasure subjects were incorrectly classified as truthful. Examiners were not able to detect successful countermeasure users.

Ginton, et al.

Only two CQT analog studies have used real, albeit small, crimes. Another study using a real crime tested the GKT technique and is discussed in the section on that technique. An experiment by Ginton, et al. (62), aimed to create even more verisimilitude than usual in the analog study. An auxiliary purpose was to test the relative effectiveness of behavioral observations, global evaluations, including behavioral observations, and numerical scoring based on the charts alone.

Subjects in Ginton, et al.'s, investigation were 21 Israeli policemen. They were given paper and pencil tests that were presented as required aptitude tests. Subjects were asked to score their own tests, which provided an opportunity to cheat, i.e., to revise their initial answers. The test answer sheets, however, were chemically treated so that cheating could be detected. Seven of the twenty-one subjects actually changed their initial answers. Later, subjects were told they were suspected of cheating, were offered an opportunity to take a polygraph examination, and were told their careers might depend on the outcome. Fifteen subjects actually underwent the polygraph testing, only two of whom had actually cheated.

A CQT was administered, and each subject was evaluated by three polygraph experts who had conducted or witnessed the particular examination being evaluated. One examiner (an observer) relied on behavioral observation, another (a rater) used only the charts, and a third (the actual examiner) used both sources of information. The evaluations were made globally. Five other polygraph examiners evaluated the charts later using both the Utah group's scoring system (21) and global evaluations. The original three performed a second analysis in this way, too. Conclusions about this study are limited because of a large no-show rate among the guilty subjects. Both guilty subjects who took the test were correctly detected. However 15 percent of the noncheaters were incorrectly identified as deceptive.

Heckel, et al.

Another analog study (74) used a staged crime to investigate the differential accuracy of CQT with psychotic, neurotic, and normal subjects. Fifteen subjects (five from each of the above three groups) were given the opportunity to steal money from the wallet of an experimenter who was staging a session of psychological testing. The experimenter later alleged that $20 had been stolen, and arranged for polygraph examinations of the 15 subjects by a field examiner. No money had actually been stolen, so the subjects were actually innocent. Four polygraph experts later rated the charts. Averaging the results for these independent evaluators, 11 of the subjects were correctly labeled innocent, 1 was called guilty, and 3 were placed in an inconclusive category. The one error and one inconclusive were with psychotic subjects, and the other two inconclusives were with neurotic subjects. Because only innocent subjects were included, a lambda was not calculated for this study.

Hammond

Hammond (64a) conducted a mock crime study to test the hypotheses that: 1) alcoholics would be less detectable than normal subjects, 2) psychopaths would be as detectable as normal subjects, and 3) student examiners would not be as accurate as an expert examiner. He was also interested in the overall value of polygraph examinations for forensic psychology. The subjects in Hammond's study were volunteers solicited through sign-up sheets in a college fraternity (normals), alcoholism treatment centers (alcoholics), and ex-offender programs (psychopaths) as well as through newspaper advertisements and other means. Psychological tests (e.g., subscales of the MMPI) as well as polygraph examinations were given to the subjects. The polygraph examinations were conducted by students near the end of their training at the Backster School of Lie Detection. Examiners used a version of Backster's control question technique, and Backster's numerical scoring system. Charts were scored using several levels of inconclusive zone by both the student examiners and an expert examiner who scored the charts blindly. Two polygraph charts, rather than the standared three, were conducted for each subject.

Table 6 shows the results of Hammond's study using the standard ±8 inconclusive zone. As shown, approximately 72 percent of the guilty subjects and 40 percent of the innocent subjects were scored correctly. Neither alcoholics, normals, nor psychopaths showed differences in detectability. In addition, there were no differences between the numerical scores of the student examiners and the blind expert examiner. However, using the ±8 cutoff, expert evaluators had more inconclusives (and fewer innocent "hits") than the student examiners. While Hammond concluded that his study supported the validity of polygraph testing, he believed that certain factors in his study could account for the failure to show differences by subject category. In particular, all subject groups actually turned out to be relatively heavy drinkers. Hammond also contended that overall accuracy rates would have been higher with more experienced polygraph examiners. He observed that the examiners in his study were unskilled at detecting countermeasures and at calibrating the polygraph instrument.

Szucko and Kleinmuntz

A somewhat different approach to assessing the validity of the polygraph was taken by Szucko and Kleinmuntz (160). They directly compared the ability of polygraph examiners to assess deception against the ability of computers to do the same using a digitalized form of the same data. They had a sample of 30 psychology undergraduate volunteers and randomly assigned them to the guilty or innocent conditions. The mock crime involved the "theft" of a $5 bill. Polygraph tests were administered by four examiner-trainees from a polygraph firm near Szucko and Kleinmuntz's university. The recordings of the physiological measures were transformed into digital form for computer analysis.

Six experienced polygraph examiners independently evaluated the charts. No inconclusive category was allowed in the study. Digital polygraph data was evaluated by computer. A lens model equation drawn from studies of human judgment was used. The results of this analysis indicated that five of the six polygraph raters were able to detect deception significantly better than chance, but four of them also had fairly high rates of false positives. Szucko and Kleinmuntz estimate that the judges detected on the average 71 percent of guilty subjects, but also called half of the innocent subjects deceptive (false positive). Szucko and Kleinmuntz state that 80 percent of the protocols could be classified correctly using a purely statistical analysis, but they do not state the detection rate, false positive rate, and false negative rate of their statistical analysis.

Kircher and Raskin (91) contend on the other hand that evaluators using numerical evaluations can be "at least as accurate as those produced by any known statistical decision model and that the accuracies of both clinical and statistical methods exceed 90 percent." Kircher and Raskin reanalyzed charts from the Rovner, et al. (143), study described above and used a lens model, similar to that employed by Szucko and Kleinmuntz. The studies, however, differed in a number of ways, which could account for the variation in their results. Probably the most important difference is that Kircher and Raskin's interpreters were trained in numerical scoring procedures (136), whereas interpreters in the Szucko and Kleinmuntz study used global evaluation procedures (139).

CONCEALED INFORMATION TESTS

Although the largest number of analog studies investigate CQT, several analog studies have examined the validity of the guilty knowledge test, one type of concealed information test. A search of the literature revealed no analog studies of the peak of tension test as a distinct technique.

Lykken

In one early investigation of GKT, Lykken (105) attempted to demonstrate that the detection of incriminating knowledge about a crime can be done more accurately than the detection of a lie about the crime. In Lykken's study, 49 male college students were randomly assigned to four categories of guilt in conducting two mock crimes. Subjects either committed a staged "theft," a staged "murder," both, or neither. An experimenter then conducted two GKT polygraph examinations with each subject, one for each crime.

Each test in Lykken's study (105) included six questions about details related to the "murder" situation and "theft" situation (e.g., asking the subject to identify an object present in the "murder" room). To make subjects anxious about the accuracy of their responses during the examination, they were told they would be given an electric shock if the examiner felt their responses indicated guilt; in fact, subjects received an electric shock after every question. The relevant alternative in each question was randomly varied among an average of five possibilities. If the question about the relevant detail produced the EDR with the greatest amplitude, it received a score of "2." If it was the second largest in amplitude, it received a "1." A perfect guilty score on each test was "12," and a perfect innocent score was "0." A score of seven or greater was categorized as guilty for the purpose of analysis, and a score of six or less was categorized as innocent. The guilty knowledge test was accurate to a significant degree in identifying subjects who committed both, either, or neither of the crimes. On the basis of this experiment, Lykken argued that GKT, with some refinements, could be applicable in criminal investigations.

Davidson

Other researchers have used Lykken's GKT paradigm to explore further its validity as a polygraph examination technique. Davidson (46) examined the GKT's validity under conditions that varied motivation level and that he claimed were, in general, more "ego-involving" for subjects. In Dawson's study, 48 college students were recruited and assigned randomly to 12 groups of 4. Three of the four were instructed to attempt to commit specific mock murders, and the fourth served as a control. The mock crimes were arranged such that one subject would "commit" the crime, one would try to fail, one was motivated but never had the opportunity, and one (the control) had no knowledge of the crime. Half of the subjects who "committed" the murders received a large amount of money ($25 to $50) and half received a small sum ($10 to $1). The different amounts were presumed to create a different level of motivation in the subjects. The subjects then were examined with the use of GKT. Six multiple-choice questions with five alternatives were presented to the subjects, and the EDR was recorded. The scoring method followed Lykken's (105) exactly (see above). Using a weighted average, 98 percent of the classifications were correct against a chance level of 25 percent. The only error was one false negative.

Podlesny and Raskin

Podlesny and Raskin (127) included GKT in their study of a variety of polygraph techniques and physiological measures. Their experiment was unique in that it employed GKT in the same context as CQT (see above). Thus, they were able to compare the accuracy rates of the two techniques, although they claimed that a different statistical comparison was impossible because the two techniques use very different methods to assess guilt. Podlesny and Raskin also were the first to test GKT with physiological measures other than EDR. To make assessments of guilt, they used the traditional polygraph respiration and cardio measures, and another vascular measure that was a composite of finger blood volume and finger blood amplitude. This latter measure was recorded by the photoplethysmograph mentioned above. In addition, Podlesny and Raskin performed a quantitative analysis of differences between guilty and innocent subjects on several other physiological measures.

GKT was conducted after the same mock theft Podlesny and Raskin (127) used to study CQT. Twenty subjects (10 guilty and 10 innocent) were examined with GKT, which included five questions with six alternatives each. The relevant alternatives were placed among the other alternatives in a "pseudo-random" order (127). The GKT charts were scored by the same method used by Lykken (105) and Davidson (46). Podlesny and Raskin also scored the charts in another way, with the addition of an inconclusive zone of scores five or six. This scoring system for assessing guilt was used with the photoplethysmograph, respiration, and cardio measure as well as EDR. Their findings were that GKT with EDR was correct for 90 percent of the subjects and incorrect for 10 percent, all false negatives. Using an inconclusive zone did not add significantly to the accuracy of the technique, however: 80 percent of assessments were correct, 10 percent incorrect (all false negatives), and 10 percent inconclusive.

Giesen and Rollison

Giesen and Rollison (61) studied the effects on GKT of the subjects' trait anxiety levels and of the possibility that crime-related details could be relevant to innocent subjects because of associations unrelated to the crime. Trait anxiety is anxiety that is characteristic of one's personality and would be relatively stable over time. Both trait anxiety and "innocent associations" could conceivably confound the detection of guilt with GKT.

Giesen and Rollison selected 40 female undergraduates who responded positively to a questionnaire item on "palmar sweating." EDR is related to sweating. Thus, this sample may have tended to produce higher EDRs than the norm. This group was divided into two groups of 20: those who scored high on a questionnaire measure of anxiety (Lykken's activity preference questionnaire) and those who scored low. Ten subjects in each group were then assigned to the guilty knowledge condition, and to the "innocent associations" condition. The guilty subjects were told to pretend to be secret agents who had committed a murder. They read a narrative about the crime, and role-played the act of burning an incriminating picture. Innocent subjects also played secret agents, but read a narrative containing several details (e.g., how much money was involved), which in the guilty condition were related to the crime. They had, therefore, as much exposure to this information as the guilty subjects, but in an innocent context. Using GKT with EDR, experimenters asked subjects eight crime-related questions, each with five alternatives. Those details common to both conditions were used as the crime-relevant items in GKT questions. Scoring followed Lykken's (105) method.

Giesen and Rollison found that GKT was highly accurate, correctly classifying all of the innocent subjects and detecting all but one of the guilty subjects (an average of 97.5 percent correct). In addition, they found that the EDR measure was significantly different between guilty and innocent subjects. Trait anxiety level had no effect on EDR by itself, but the more anxious subjects in the guilty condition had significantly greater EDR than the less anxious, especially in response to the relevant items. These findings would suggest that anxiety alone does not confound GKT results, but anxiety in guilty subjects might indeed augment the accuracy of the technique. The study also suggests that GKT may be accurate even when innocent subjects have greater associations with crime-relevant items than with neutral items. This finding, however, must be tempered by the fact that the entire sample was selected for their tendency for palmar sweating under stress and, thus, may be unrepresentative of polygraph subjects in general.

Balloun and Holmes

Balloun and Holmes (12) used GKT to detect guilt in a "real" crime arranged by the experimenters. They were also interested in the effect of psychopathy and of repeated examinations on the accuracy of GKT. They selected 18 male college students with high scores on the psychopathic deviate (Pd) scale of the Minnesota Multiphasic Personality Inventory (MMPI) and 16 with low scores. The Pd scale was originally designed to make the diagnosis of psychopathic personality and was used as a scale to measure relative "amounts" of psychopathy. The experimenters acknowledge, however, that the Pd scale may be an inadequate measure of this diagnosis. These subjects took a fake intelligence test with two other students (actually confederates of the examiner). The confederates urged subjects to cheat and supplied test answers to those who were willing. Eighteen of the thirty four students cheated. Later, the subjects underwent a polygraph examination using GKT. They were reminded that cheating on exams could lead to academic dismissal, and that the experimenters knew that some had cheated on the "intelligence test." Information from the intelligence tests that only the cheaters would know served as the incriminating details on GKT. Another GKT with the same content, but a different order of questions was then administered to see if the subjects would adapt to GKT and, thus, reduce its accuracy.

Balloun and Holmes scored GKT using Lykken's (105) method with three physiological measures (EDR, heart rate, and finger pulse volume), but only EDR produced significant results. On the first test, guilty subjects scored significantly higher and were detected with significant accuracy. However, on the second test, though the guilty subjects had significantly greater scores, they were not great enough for significantly accurate detection of guilt at the criterion level (5.5 out of 10) used. There was no difference between the high and low Pd subjects on either administration of GKT.

Bradley and Janisse

In their study of the influence of threat and demonstrations of accuracy on the polygraph examination (see above), Bradley and Janisse (35) also tested the 192 subjects with the GKT after the CQT had been conducted. The questions concerned four relevant details. They were scored using the Lykken (105) method. With EDR data, the GKT classified an average of 74 percent of subjects correctly, and 26 percent incorrectly with 11 false positives and 39 false negatives. With the measure of heart rate change, the GKT categorized 63.5 percent of subjects correctly and 36.5 incorrectly, with 17 false positives and 53 false negatives. Neither the degree of threat nor the demonstrated effectiveness of the polygraph test had a significant effect on the discrimination between deceptive and truthful subjects.

Timm

Timm (163) examined the effect of the administration of a placebo on the validity of GKT. Also included in the experiment was an investigation of the effect on GKT accuracy of differential feedback from the stimulation test. In the experiment all 270 college student subjects committed a mock crime. There were no "innocent" subjects. Before the mock crime, subjects were either: 1) given a placebo and told it would help them "beat" the test; 2) given a placebo and told it would make it more difficult to deceive the examiner; or 3) not given a placebo. The stimulation or number test was arranged to produce three different feedback conditions. One-third of the subjects' numbers were detected, one-third were not, and one-third did not receive the results of the stimulation test. After the GKT was conducted on each subject, charts were scored according to the Lykken (105) method. Adequate charts were obtained for 237 subjects. Of these subjects, 70.4 to 80.8 percent of them produced scores indicative of guilt, depending on how conservative a cutoff point for the score was used. Neither the placebo condition nor the feedback condition produced a significant effect on detection ability. Because of the absence of "innocent" subjects in this study (i.e., a base rate of guilty of 100 percent), the study tells us nothing about the accuracy of GKT with the innocent subjects. And even the results with guilty subjects are difficult to interpret when there is no comparison to results with innocent subjects. Also, without innocent subjects, a lambda is impossible to calculate.

PREEMPLOYMENT SCREENING

Despite its widespread use in the field, there are few analog studies of the preemployment screening polygraph examination. The two that are known to employ post-1960 polygraph screening techniques are reviewed. Correa and Adams (43) conducted an analog investigation of this type of examination with 40 undergraduate subjects. Barland (16) conducted an analog study with Federal Government personnel.

Correa and Adams

Like the usual preemployment screening test, the examination in Correa and Adams' study included a number of relevant questions. Subjects were interviewed prior to the polygraph examination and completed a questionnaire about their background. Half the group was instructed to lie to nine relevant questions and half to tell the truth. The polygraph test was conducted, and three charts of 32 questions each were recorded. Most of the relevant questions concerned information from the questionnaire, but also included were three questions about events staged by the researcher in the initial interview (e.g., giving the subject a glass of water). These latter questions served as a check on the honesty of subjects in completing the questionnaire, and were considered relevant questions in the evaluation of deception or nondeception. The examiner subjectively made assessments of veracity based on the polygraph recordings. When questions about the staged events and the application were diagnosed by the examiner, all 40 of the subjects were correctly identified as being deceptive or truthful.

Correa and Adams conducted a question-by-question analysis of the charts of deceptive subjects. A mean of 75 percent of the relevant items from the screening application were correctly classified, and a mean of 25 percent were incorrectly classified. When change scores were calculated for each physiological response, all physiological measures (EDR, respiration, cardiovascular) significantly discriminated truthful from deceptive subjects. Correa and Adams suggest that these findings provide evidence for the validity of prescreening polygraph examinations. There are, however, a number of problems with the Correa and Adams' study that may compromise its validity. Several features of the experiment are probably highly unrepresentative of or unrelated to field preemployment polygraph examinations: the length of the interview (96 questions); the number of deceptive responses subjects made (9); and the inclusion of questions about the experiment itself. Furthermore, the experimenters fail to discuss the criteria by which the assessments of veracity were made, so it is difficult to ascertain whether these assessments correspond to field assessments.

Barland

The Barland (16) study is important for several reasons. One, subjects were actual military personnel who in Barland's opinion might be the types screened for counterintelligence purposes. All subjects were assigned to intelligence duties. It is, thus, unique in being the only validity study of preemployment screening in an intelligence context. However, because it did not ask any questions related to security interests, it cannot be considered a full analog to field situations.

Second, it tested the validity of a type of CQT, the directed lie control question (DLCQ) technique, in a screening situation. DLCQ is part of a counterintelligence screening test developed by Army Intelligence examiners in 1971. During the pretest phase of this technique, subjects typically answer "yes" to certain questions. When they answer yes, the examiner instructs them that when they are asked such questions during the actual polygraph examination, they should respond with a "no" rather than a "yes." Thus, they are directed to lie, and their lies to these questions constitute the control questions against which responses to relevant questions are compared. DLCQ differs from the control question discussed previously (see ch. 2). With the DLCQ technique, the control questions are not designed to provoke the subject to lie or be concerned about the telling the truth. The "lies" do not constitute deception since the examiner instructs the subject to tell lies that they both know are false. However, the directed lies are believed to generate concern in innocent subjects because the subjects are told that to appear nondeceptive on the rest of the examination, they must appear deceptive on the directed questions.

The question of whether CQT can be used outside of specific issue investigations (e.g., in preemployment or periodic screening) is controversial. It is difficult to construct standard control questions when much of a person's past is irrelevant to the purpose of the examination, since past misdeeds (i.e., other than the specific issue being investigated) typically comprise the subject area of control questions.

In this 1981 study, Barland solicited volunteers from the military intelligence community. Subjects were told the purpose of the study and that testing would be limited to the subject's date of birth, place of birth, education, employment, and residences (these were the relevant items), and that some subjects would be instructed to furnish the examiner with false information. Approximately half the subjects were instructed to lie to one of the relevant items; these subjects were offered a $20 reward if they could appear truthful on the polygraph examination. Unlike the data in the Correa and Adams' study, the experimenter was able to check the information given by the subjects against data obtained from background investigations. The three polygraph examiners in the study had 3, 6, and 9 years of polygraph experience and had been trained at the U.S. Army Military Police School (USAMPS) polygraph course.

Examiners used three methods of chart interpretation: zone of comparison, greatest control method, and relevant-irrelevant method. As explained in chapter 2, in the zone method, relevant questions are evaluated against the larger of either control question response in a zone. In Barland's (16) zone method, each physiological measure for each relevant/irrelevant control question pair was rated on a point scale using interpretive criteria taught at USAMPS. In the relevant-irrelevant method of interpretation, each relevant question was evaluated without making specific reference to the control question nearest it; emphasis "was placed on the size and consistency of reactions at the relevant questions" and scored globally rather than numerically. The "greatest control" method consisted of evaluating all five relevant questions against the single control question on that chart which had the largest overall reaction. In addition to the comparisons of the three chart interpretation methods, charts were analyzed globally and on a question-by-question basis.

In the global method of analysis, subjects were categorized as either deception indicated, no deception indicated, or inconclusive on the basis of appearing deceptive to any of the relevant questions. That is, if a subject was in fact deceptive to any relevant question, and he reacted deceptively to any of the questions, it was considered a hit even though the examiner may have misidentified which relevant question the subject was deceptive to. Using this method of assessing deceptiveness, the three methods of chart interpretation achieved the following results:

Zone:
  - 62 percent correct identification of truthful subjects;
  - 19 percent incorrect;
  - 19 percent inconclusive;

  - 70 percent correct identification of deceptive subjects;
  - 17 percent incorrect;
  - 13 percent inconclusive.

Greatest control:
  - 77 percent correct identification of truthful subjects;
  - 15 percent incorrect;
  - 8 percent inconclusive.

  - 50 percent correct identification of deceptive subjects;
  - 23 percent incorrect;
  - 27 percent inconclusive.

Relevant-irrelevant:
  - 73 percent correct identification of truthful subjects;
  - 23 percent incorrect;
  - 4 percent inconclusive.

  - 80 percent correct identification of deceptive subjects;
  - 13 percent incorrect;
  - 7 percent inconclusive.

Presumably, the correct identification rates would be lower if only those cases in which the truly deceptive relevant response was counted as a "hit." To test this hypothesis, the authors conducted a question-by-question analysis. In this method, identification of truthful responses increased but identification of deceptive responses declined quite a bit. Using the zone technique, 77 percent of the truthful questions and only 57 percent of the deceptive questions were correctly identified. With the greatest control scoring method, 85 percent of truthful responses and less than half (43 percent) of deceptive questions were correctly identified. The R/I scoring technique showed the best results. With this method, 88 percent of the truthful subjects and 67 percent of deceptive questions were correctly identified (although global results were better with the R/I technique). This interpretation should be modified by the fact that each examiner used all three scoring techniques and the R/I technique was the last one used. Thus, the interpreter had the benefit of his previous judgments. The results of a blind analysis using other interpreters were not ready to be reported by Barland at the time his 1981 report was submitted.

The results of the Barland study raise serious questions about the usefulness of directed lie control questions in screening procedures as well as, in general, the validity of polygraph testing for preemployment and counterintelligence purposes, especially if used alone. Of course, the limitations of analog studies should be taken into consideration. Because of these limitations, Barland considers his results a "worst case" scenario. Finally, interpretations must depend on the false positive and false negative rates which are deemed acceptable for particular purposes.

FINDINGS

Separate statistical analyses were performed for the guilty knowledge and control question analog studies. The following data for the analog studies discussed above were reviewed:

- percentage of guilty subjects judged deceptive;
- percentage of guilty subjects judged nondeceptive (false negatives);
- percentage of guilty subjects judged inconclusive;
- percentage of innocent subjects judged deceptive (false positives);
- percentage of innocent subjects judged truthful; and
- percentage of innocent subjects judged inconclusive.

Also, as with the field studies, an index of predictive association (lambda$_b$) was calculated. The results (see tables 8 and 9) indicate that the control question test provides a 43-percent improvement in prediction over the base rate for these analog studies, and the guilty knowledge test a 70-percent improvement in prediction over the base rates. Because the studies differed so much, lambdas were calculated separately for each study. As shown in tables 6 and 7, individual lambdas ranged from zero to 83 percent for the CQT studies and 38 to 95 percent for the GKT studies (see ch. 4). These figures should be interpreted with caution as in real life the base rate of guilt will vary considerably from approximately 50/50 distributions in laboratory experiments. Thus, it is difficult to draw unqualified conclusions from the analog studies given the wide variety of designs used.

The false negative rate for the analog studies of CQT technique ranged from 0 to 29 percent. Inconclusives ranged from 0 to 44 percent for guilty subjects and from 0 to 53 percent for innocent subjects. There is a wide range of false positives (4 to 51 percent). Global evaluations by the examiners, field scoring techniques, and purely statistical analyses of the data all seem to produce high detection rates in most studies. One exception is Kleinmuntz and Szucko's (92) study, which found the validity coefficients of polygraph examiners' judgments markedly inferior to a purely statistical analysis of the charts. However, it is unclear how comparable their method of measuring validity is to the usual method of using an accuracy rate, and it is also not clear how applicable the lens model they use is to the question of the validity of the polygraph.

Another exception is Ginton, et al.'s, study (62), in which field numerical scoring was found to be inferior to the global evaluation method in detecting deception. However, the examiners in that study were Israeli polygraph professionals who may characteristically use a global method of assessment, and who may have been unfamiliar with the Utah numerical scoring system.

Accuracy of detection differed sizably between control question analog studies using students as subjects (Barland and Raskin, Bradley and Janisse, Szucko and Kleinmuntz; Widacki and Horvath is excluded as discussed above) and other control question analog studies (Podlesny and Raskin, Raskin and Hare, Rovner, et al., Dawson, Ginton, et al.). Experiments using students had lower percentages of correct decisions for both guilty and innocent, and more false negatives and false positives. Given the small number of studies in each category when the studies are divided in this way, it is unclear whether this difference is attributable to the nature of the subjects (student v. nonstudent) or other characteristics of these experiments.

As shown in tables 8 and 9, GKT analog studies detected a slightly lower average percentage of the guilty subjects than the CQT analog studies. They also had a relatively higher proportion of false negatives but a lower rate of false positives. It should be noted, however, that GKT was not assessed under conditions that deviated as much from the ideal as the control question test deviated. Nor were there as many studies testing GKT as CQT. This suggests that the confidence one can have in the GKT findings is, in general, less than the confidence one can have in the CQT findings.

In summary, there exists a number of studies of CQT; a smaller number of the concealed information test, all using GKT; and only two studies of the preemployment screening interview, one of them with Government personnel. The analog studies systematically explored many of the technical variables associated with the polygraph (cf. the Utah group's studies of CQT), and also studied the effect of several situational variables on the validity of the polygraph. The control question test was found to detect guilty subjects with a relatively high degree of accuracy, but also to be subject to false positive errors. There was a large amount of variability among the control question analogs, especially the more they diverged in technique from the field method. The guilty knowledge test had a slightly lower rate of detection of guilt, more false negatives, but fewer false positives.

Chapter 6

Factors Affecting Polygraph Examination Validity

Chapter 6

Factors Affecting Polygraph Examination Validity

INTRODUCTION

The analyses of both field and analog studies reported in chapters 4 and 5 indicate that there is considerable variability in accuracy rates of polygraph examinations. To interpret these variations, numerous factors, such as the restricted range of techniques and applications tested in these studies, need to be considered. In addition, researchers have attempted to explain the variability in accuracy scores by proposing a number of factors that theoretically may affect polygraph test validity. These include characteristics of examiners, settings, and subjects. In addition, subjects have been known to use, or might be trained to use, a number of countermeasures to "beat" the polygraph. For many of these factors the research evidence is contradictory. For others, there has been little or no empirical testing. This chapter describes evidence from field and analog studies, as well as from laboratory investigations, on factors that may affect the accuracy of polygraph tests. The chapter also discusses possible priorities for additional research on factors affecting polygraph validity.

POLYGRAPH EXAMINER, SUBJECT, AND SETTING

The previously described analyses of field and analog studies (see chs. 4 and 5) emphasize the characteristics of polygraph tests and their relation to accurate or inaccurate outcomes. In the present section, the focus shifts away from the tests themselves, to additional factors that may affect validity. These factors are sometimes referred to as dimensions of external validity and aid in the assessment of the generalizability of research findings. Considerations of these factors will enable evaluation of the conditions under which various levels of validity may be expected from polygraph examinations. Differential validity in polygraph tests may be obtained with different examiners, subject populations, and with examinations conducted in different settings.

Examiner

It has long been recognized (cf. 108,122,135, 154) that the examiner's skill has an important effect on the validity of polygraph tests. Examiner experience is an essential element reported by investigators and has often been used to explain differences in accuracy rates (137,138). There are some data to indicate that experienced examiners have better accuracy rates. In recognition of this outcome, training has been accorded a high priority both within and outside Government agencies which conduct polygraph examinations and by polygraph examiner groups (cf. 3). An extensive array of training facilities now exists, offering a somewhat diverse set of orientations to polygraph testing.

Experience

A number of studies have tested how examiner experience relates to validity of polygraph examinations. Horvath and Reid (84), for example, had charts utilized in their validity study reexamined by a group of 10 polygraph examiners. Seven of the examiners were experienced and three of them were examiner-interns (each with less than 6 months' experience). According to Horvath and Reid, experienced examiners made an average of 91.4 percent correct judgments, while the average for inexperienced examiners was 77.5 percent.

Training

Experience in conducting polygraph examinations suggests that there are a number of clinical components to detection of deception. To some extent, training programs capture these clinical elements by extensive training in "proper" examiner attitude and relationship with subjects. Increasingly, however, training programs emphasize standardized techniques for constructing questions and scoring examinations. In this respect, the U.S. Army Military Police School (USAMPS) is perhaps the best example. The school serves as the central training site for almost all Government agencies which maintain polygraph examiner staffs. USAMPS teaches several versions of the control question technique (CQT) (including what they call the modified general question technique (MGQT) and the original Backster's zone of comparison (ZOC) method) and several specific protocols for selecting question sets and scoring polygraph charts. Trainees receive both didactic classroom training and supervised experience conducting polygraph examinations. The current curriculum for USAMPS uses Reid and Inbau's (139) text on polygraph testing, supplemented by materials prepared especially for its trainees (179). USAMPS is one of a number of training programs certified by the American Polygraph Association (cf. 3).

On the basis of presently available data, it is not possible to determine whether types of training have an effect on outcomes. A study by Raskin (133) indicates that examiners trained in schools that emphasize numerical scoring were significantly more accurate than examiners who attended other schools (97.1 v. 86.9 percent). It is difficult to determine, however, if training in numerical scoring is more efficient or if better examiners/schools select such techniques. The fact that examiners who were trained in numerical techniques, but who did not use them, did more poorly than examiners trained in numerical techniques who used them (88.5 v. 98.9 percent) suggests that numerical evaluation rather than examiner selection (or some other aspect of the training) provides an advantage.

Subjects

Much effort in recent years has been devoted to development of systematic training. Less attention appears to have been paid to the characteristics of subjects of polygraph testing. Frequently, research reports of polygraph examination do not report even the most easily available data on subject characteristics (e.g., proportion of males and females). There have, however, been a number of studies of specific population groups (e.g., psychopaths) hypothesized to be less detectable. In addition to subjects' psychopathy, other diagnostic categories and subject variables such as gender, intelligence, motivation, and responsivity to arousal may also affect validity.

Subject factors are often described in the literature as personality or individual difference factors (136,194). They refer to traits associated with individuals that may make them differentially detectable in a polygraph examination. Understanding these effects should enable determination of the conditions under which polygraph testing will yield particular levels of validity. The mechanism by which subject variables affect polygraph examination validity has to do with differential autonomic arousal. Validity is affected when an interaction results between arousal and polygraph testing.

Psychopathy and Level of Socialization

One aspect of potential subject effects that has received considerable attention is the effect of level of socialization and psychopathy on detectability. In a series of studies by Waid and his colleagues (193,198,199) significant relationships were found in the laboratory between socialization and autonomic responsiveness. An initial finding (193) was that college students who scored low on socialization (on a standard psychological inventory), gave smaller electrodermal responses (EDRs) to stimuli than did high scoring subjects. In a more directly relevant investigation (198), a group of college students was asked to deceive or not to deceive a professional polygraph examiner. Results indicated that subjects who were not detectable were significantly less socialized than those who were detectable. Susceptibility to detection seemed to be mediated by socialization; results indicated that low socialization subjects showed reduced EDRs. Highly socialized subjects were more responsive electrodermally, and as a result, several of them were misclassified as deceptive.

Raskin (136) has criticized Waid, et al.'s (198), research as not having practical significance for evaluations of polygraph validity. According to Raskin, simply demonstrating that there is a difference in responsivity on the first set of questions does not mean that subjects would not be correctly detected in an actual polygraph examination (which may involve three to four charts). Some of Raskin's own studies (e.g., 21,137) suggest that psychopathic individuals are not less detectable than nonpsychopathic individuals. In Raskin and Hare's study, convicted felons, half of whom were diagnosed as psychopathic, performed a mock crime. These subjects were then administered a polygraph examination and offered a substantial monetary bonus if they could produce a truthful outcome. In contrast to Waid, et al.'s, findings, Raskin and Hare found that individuals diagnosed as psychopathic and/or low in socialization were more reactive and easily detectable than those not psychopathic and high in socialization. Earlier research by Raskin (21) supports this finding. Barland and Raskin's (22) field study, on the other hand, found that subjects who scored high on the psychopathic deviate (Pd) scale of the Minnesota Multiphasic Personality Inventory (MMPI) (a measure of psychopathy) had smaller cardio (but not respiration or skin conductance) scores than low Pd subjects.

In a previously described study, Balloun and Holmes (12) conducted an analog study of college students using a "cheating" situation. Their results indicated that subjects who scored high on the Pd scale of the MMPI were just as easy to detect as were those individuals who scored low on the scale. It is important to note, however, that the polygraph test was a concealed information type of test, not a CQT or relevant/irrelevant (R/I) test. A doctoral dissertation by Hammond (64a) also found no differences between normal and psychopaths.

Other Psychopathology

Guilty psychopaths may escape detection because they are not concerned enough about a misdeed to create interpretable physiological responses. Individuals with other forms of psychopathology may escape detection or be classified as false positives for other reasons (e.g., emotional instability, delusional thinking). The one study that has investigated this possibility (74) found, in fact, that innocent neurotics and particularly psychotics were likely to be identified as deceptive. There were no guilty subjects in this "real crime" analog study.

Gender

One of the most obvious subject differences is gender. Males and females may have different patterns of autonomic arousal, and such differences may affect polygraph testing validity (136,194). Unfortunately, few data exist to examine this hypothesis; most research only studies male subjects. The one study by Cutrow, et. al. (45), that specifically tested for sex differences did not find any. In another study (61), all female subjects were tested in a mock-crime situation using the guilty knowledge test (GKT). GKT was found to be highly accurate, but because males were not also tested, it is impossible to determine if males would have been less detectable. The two Honts and Hodes (76,77) analog studies described in chapter 5 included both females and males; the researchers do not report any gender differences in detectability.

Intelligence

Intelligence is an additional variable which potentially might affect detectability. The ability of intelligent subjects to anticipate questions may affect polygraph accuracy. One possibility is that intelligent subjects are less detectable because, if trained, they are able to anticipate questions and employ countermeasures. Another possibility is that because intelligent subjects better understand the implications of a polygraph examination, they will respond to relevant questions with heightened arousal when they are attempting to deceive (20).

There has been relatively little research on intelligence and polygraph testing. In one of the few experiments which look at intelligence and detectability, Kugelmass (95) found no correlation between intelligence and responsivity on a peak of tension (POT) card test. On the other hand, research by Gustafson and Orne (65) found that motivation to deceive increased the probability of detection. Barland and Raskin (20) feel this is evidence of the potential role of intelligence. Barland and Raskin's study (22) which compared detection rates among subjects of different education levels, found no difference. However, a separate analysis of the sources of false positive errors by Raskin (133) found that the majority of false positives occurred among subjects who had college degrees. Level of education, of course, is not a perfect indicator of intelligence, and there is a need to better understand these relationships.

Ethnic and Group Differences

Another category of subject differences that may affect polygraph validity has to do with ethnic and group differences in physiological response. Research conducted cross culturally (e.g., 97,104,158), indicates that there are ethnic differences in response to stress. Such differences may, in turn, affect detection of deception. As noted earlier, these effects may interact with the ethnic identification of the examiner. However, effects of ethnic differences have not been directly tested with respect to polygraph examinations.

Autonomic Lability

A final individual difference is what Waid and Orne (194) have referred to as autonomic lability. Regardless of other differences among subjects, there may be consistent individual differences connected with their level of autonomic arousal.

Although there is considerable variance for an individual in autonomic responses to most physiological measures of autonomic nervous system (ANS) arousal, electrodermal lability may be different. Given the importance of the EDR for polygraph examinations, it may be essential to understand more about this factor. Unfortunately, most of this research (e.g., 200) has been conducted with concealed information tests and not with CQT or R/I tests.

Setting

One theory underlying lie detection using the polygraph is that the threat of punishment leads an individual to manifest a physiological reaction (48). This suggests, then, that settings in which an individual is more certain of being detected and in which the consequences are greatest, will permit higher levels of detection. Furthermore, in order to be certain of being detected, a subject must believe in the efficacy of the polygraph procedures in order for it to function. According to some (e.g., 194), the polygraph is often used somewhat like a "stage prop," and its presence is meant to "enhance the subject's concern." Stimulation tests, used in almost all field polygraph examinations, serve the same function, albeit more directly. There is considerable discussion (e.g., 202) in the literature about how frequently within a polygraph examination such stimulation tests should be utilized in order to increase the validity of the examination.

Instrument

Some research, reported by Orne and his colleagues, addresses the question of the situational features necessary for a polygraph examination. In one component of a study reported by Orne, et al. (123), subjects were led to believe that the polygraph recording equipment was not operative. There was some indication that the pretest condition in which subjects were led to believe that the polygraph instrument was inoperative produced a lower detectability; however, results were not statistically significant. In an earlier study (161), detectability was not affected by subjects' belief in whether the machine was recording. Both of these studies involved use of concealed information tests.

A more recent study by Orne's group (198) tested a similar hypothesis using a different procedure. In this study, subjects saw the polygraph machine turned off, although the experimenters actually ran the leads to a second polygraph device and were able to record responses during a pretest review of questions. The results indicated that subjects who were aware of being recorded had significantly higher responses to relevant questions and not significantly different responses to control questions.

Bogus Pipeline

An interesting and potentially important aspect of how the polygraph achieves valid results is based on what social psychologists such as Jones and Sigall call the "bogus pipeline" (87). The bogus pipeline is a procedure used to elicit truthful attitudes in situations where social desirability effects (i.e., subjects' desire to express socially acceptable opinions) may mask actual attitudes. The procedure involves attaching subjects (via skin electrodes) to an ostensible physiological recording device called the "electromyograph" (EMG) and providing subjects with a "steering wheel" device to record their attitudes. In a typical study (87), subjects were told that the EMG measured implicit muscle potentials and that it was an improved polygraph or "lie detector." The recording device is actually "electrical junk" (87), and the purpose of the procedure is simply to convince subjects that their actual attitudes are detectable.

Results from a number of investigations which have used the bogus pipeline procedure (e.g., 131,150) support Jones and Sigall's premise. Several studies indicate that when subjects believe that their attitudes are detectable by a physiological recording device, they more readily express their actual attitudes. Although it is difficult to know what "actual" attitudes are, higher truthfulness is assumed with the bogus pipeline because the procedure yields more socially undesirable responses than when it is not used. For example, in Sigall and Page's (150) initial experiment, they found that subjects in the bogus pipeline condition would admit to negative attitudes about "Negroes." Similar subjects in nonbogus pipeline conditions using paper-and-pencil tests would not reveal such attitudes. Later research has shown that this findings holds for attitudes toward handicapped individuals and for "confessing" to having prior knowledge about a psychological experiment.

Although the bogus pipeline research suggests that the conditions of testing (in particular, the perceived complexity and accuracy of equipment) may have important effects on polygraph subjects, it is not clear how or to what extent these effects influence the validity of the test itself. In a substantial number of criminal investigations subjects voluntarily confess after having the polygraph procedure explained or being shown the results of the examination. In personnel screening, subjects often admit to errors in their job application or past indiscretions (24,165). Most available field and analog research does not permit determination of the extent to which the polygraph is used in this way.

Specific Settings

Polygraph examinations take place in a number of settings, ranging from facilities specifically designed for this purpose to motel rooms. Specifically designed facilities usually include one-way mirrors for observation and audio recording capabilities, and are located so as to prevent interruptions during the examination. It is reasonable to assume that the setting may interact both with subject and examiner characteristics to affect the validity of polygraph tests. No research, however, directly tests the impact of different settings on polygraph testing validity.

COUNTERMEASURES

Countermeasures are deliberate techniques used by deceptive subjects to avoid detection during a polygraph examination (23,108,139,194,195). Countermeasures can range from simple physical techniques, to so-called mental countermeasures, to the use of drugs and biofeedback techniques. There is a potentially large list of such countermeasures, and there are a number of plausible, but not yet validated, techniques to avoid deception. The research on polygraph countermeasures is summarized below by type of countermeasure.

Physical

Physical measures taken by a subject during a polygraph examination are, perhaps, the most frequently discussed countermeasures used by subjects (20,108). Any physical activity which could affect physiological response is a potential problem for interpretation of a polygraph test record. There is no question that physical measures, from tensing muscles to biting the tongue, to squeezing toes, to shifting one's position can affect physiological response.

There are frequent references to the use of such measures (see e.g., 40,108). But little systematic research has been conducted to establish the impact of the use of such measures on polygraph decisions. Kubis (93) found that when subjects press their toes towards the floor they were able to reduce the probability of detection from 75 to 10 percent. A replication of this experiment, however, by More (119) found that there was no decrease in detectability caused by toe movements. In two more recent studies discussed in chapter 5, by Honts and Hodes (76,77), the efficacy of two physical countermeasures was tested in analog situations. Both studies found that countermeasures allowed subjects to "beat" the polygraph. Furthermore, there were a large percentage of inconclusives. In addition, both studies found that experienced examiners were not able to detect use of the countermeasures. A recent study by Honts, Raskin, and Kircher (78) also found that the use of physical countermeasures decreased detectability; the false negative rate for countermeasure subjects was 78 percent. However, examiners using a separate EMG analysis were able to detect the use of countermeasures 80 percent of the time.

Thus, the evidence, while limited, is that deceptive subjects who use physical countermeasures and who can distinguish nonrelevant from relevant questions (in a CQT or R/I test) can increase their chances of avoiding detection.

Drugs

In contrast to physical measures, which potentially may be detected by an observant polygraph examiner by running multiple charts or by careful comparison of particular physiological measures, the use of various pharmacological agents is probably more difficult to detect. Not only may drugs be difficult to detect by observation, but they may also not be detected by multiple polygraph tests. Some theorists have suggested that such pharmacological agents have the potential to produce incorrect or uninterpretable polygraph records.

Research on drugs' factors is only beginning to be conducted. Recent research by Waid (197) indicates that the tranquilizer, meprobamate (Miltown®), permits subjects who are being deceptive to increase their ability to avoid detection in a polygraph examination. One feature of tranquilizers such as meprobamate is that they suppress autonomic activity which may not be accompanied by any observable psychomotor differences. In Waid, et al.'s, study a GKT was used in a polygraph test. Subjects were all male and divided into three groups: 1) a tranquilizer group; 2) a placebo group; and 3) a control group. Only 3 of 11 guilty subjects who had taken meprobamate were scored as deceptive.

It should be noted that because Waid, et al.'s, investigation involved GKT, the ability to generalize from the results is limited. According to Raskin (136), a different problem would be encountered by attempts to utilize tranquilizers to defeat an examination employing CQT. The use of such drugs in a CQT polygraph examination would be more likely to yield inconclusive findings, rather than errors, because the drugs would likely result in no difference between the responses to control and relevant questions. This interpretation is supported by the recent analog study of Gatchel, et al. (59), which found that the use of propranolol, a beta-blocking drug, resulted in a 32.2-percent inconclusive rate, although the overall error rate was low. An additional finding was that examiners could not tell which subjects had used the drug. Conclusions drawn from this study must be limited by the fact that subjects were students. Other studies using college students (e.g., 76,77) have also resulted in large numbers of inconclusives.

A recent study by Iacono, et al. (86), found that ingestion of neither 10 milligrams of diazepam (Valium®) nor 20 milligrams of methylophenidate (Ritalin®) affected the accuracy of detection. Results in both active drug conditions were more accurate than when subjects ingested a placebo (a capsule containing lactose).

Research on other psychoactive drugs has not been reported in the literature, although such research is now being planned under the auspices of the National Security Agency and the Army Intelligence and Security Command. There are also no data as to the use of common drugs by actual polygraph examination subjects. Although examiners normally ask subjects to report use of medications or other drugs, blood samples or other detection means are typically not employed. It is thus difficult to assess the magnitude of drug use by subjects in previous research on the validity of polygraph testing.

In addition to drugs, there have also been reports of the use of various chemicals to confuse physiological recording (see 20). Placing antiperspirant powder, clear nail polish, or other agents on the balls of one's finger's may make EDRs less reliable. Such measures, however, should be detectable by a trained examiner.

Hypnosis/Biofeedback

There is a substantial literature in psychology about the use of hypnosis and biofeedback to alter and condition physiological responses. There is some evidence (see 146) that hypnosis, for example, induces declines in skin conductance levels. A number of investigations have attempted to show that hypnotically suggested amnesia is an effective countermeasure. Such research seems to indicate that hypnosis is not effective (see 20).

Recent research by Corcoran, Lewis, and Garver (42) has examined the effects of biofeedback training on suppressing EDR. They found that both hypnosis and biofeedback groups were able to reduce detectability after training as compared to a control group. In another study, Rovner, Raskin, and Kircher (143) reported that subjects who received extensive information about the nature of lie detection and practiced using countermeasures were detected significantly less than subjects without such training. It seems clear that if hypnosis or biofeedback operate as countermeasures, especially with commonly used tests such as CQT, that extensive training would have to accompany their use.

Mental

Another category of countermeasures involves those that get the subject to think differently about the test. As noted earlier, most polygraph examinations rely on the subject's motivation to avoid detection rather than on any response directly connected with "telling a lie." Simple cognitive countermeasures include patterns of thinking that suppress responses to control or irrelevant questions. More complex cognitive countermeasures are based on knowledge of the results of the examination and lack of belief in one's detectability.

Controlling Thoughts

Any individual who understands the basic structure of a particular polygraph examination should be able to differentiate irrelevant and control from relevant questions (when an R/I or CQT examination is conducted). Given that when a CQT is employed, the questions are reviewed prior to their presentation, a subject intent on deception could try to alter cognitively responses to various questions (although, since the order of presentation is varied, this is made difficult during an actual test). The procedure would be to try to dissociate oneself from the relevant questions and heighten response to control questions. Various means of such mental dissociation have been described (cf. 20,108). They range from concentrating on an irrelevant object or by convincing oneself that the question means something different than intended.

There is little research evidence available to ascertain the potential of these techniques. Only one study on the use of such cognitive countermeasures appears to have been conducted, and it does not provide much information. In this study, Dawson (49) used as subjects actors trained in the "method" school of acting. Subjects were instructed to try to defeat the polygraph in a CQT examination. The results indicated that without exception the actors were detected and were not able to control their affective responses. It should be noted, however, that the actors in Dawson's study were merely told to "appear innocent" and were not told what would make them appear nondeceptive on a polygraph test.

Knowledge of Results

Another set of countermeasures is based on subjects having knowledge of the results of their polygraph examination. In criminal situations, particularly when an inconclusive outcome is obtained or when the subject disagrees with the outcome of an initial polygraph test, subjects are retested. In noncriminal situations, individuals are often tested at the beginning of their employment and at a number of subsequent points. In addition, subjects who know that they will receive a polygraph examination may seek training in methods to avoid detection. For all of these reasons, it is important to understand how feedback about polygraph examinations affects validity and whether prior experiences represent a potential countermeasure.

In an early laboratory study (67), subjects were given a stimulation test and feedback concerning its outcome. Feedback was manipulated so that some subjects thought they had successfully avoided detection and others thought that they had not. Subjects were motivated on a subsequent trial to avoid detection (they were told, "only mature and stable individuals are able to fool the lie detector"). The results indicated that subjects who believed that they had avoided detection were much less detectable on the second trial (13 out of 16 were not detected, while only 1 of 16 in a control group were able to avoid detection). It should be noted that a stimulation test is a form of a concealed information test and the result may be due to lowered overall arousal. "Beating" a CQT represents a somewhat different problem.

A recent study, by Rovner, et al. (143), tested a similar hypothesis in a CQT examination. Several groups of subjects were placed in a mock crime situation. One group was given information about the nature of a CQT examination and information on what physiological reactions they should try to simulate. Another group was given information plus two practice tests involving actual physiological recordings after which they were told whether or not they had beat the polygraph. A third group served as a control and was given a typical polygraph examination. The results indicated that the information only and control group were not able to avoid detection; however, 25 percent of the guilty subjects in the information plus practice group were able to avoid detection. Raskin (136) maintains that this 25-percent error rate should be considered the "upper limit" because, in actual field situations, motivation would be much higher. Although Raskin is, perhaps, correct, it is also possible that in actual situations (where motivation is high), subjects might engage in more practice.

Belief in "Machine"

A final countermeasure is based on research about the bogus pipeline (87) and the role of the setting in inducing valid outcomes. If the validity of polygraph testing is dependent on the belief by subjects in the efficacy of the procedure, then a possible countermeasure would involve training subjects to believe that the polygraph does not work. This might be done, for example, by providing subjects with false feedback on a polygraph examination. Unfortunately there is little research in this area, and the two studies that have been conducted come to different conclusions about the effect of belief in the techniques' effectiveness. In one study, Bradley and Janisse (35) tested the hypothesis by rigging a stimulation test at various levels of effective detection. Depending on the condition, subjects were "detected" on one, two, or three trials to create the impression that the detection measures were ineffective, sometimes effective, or perfectly effective. For the EDR measure, the more effective the apparatus appeared to be, the more the innocent subjects scored as nondeceptive and the more the guilty subjects scored as deceptive. In an earlier study, however, Timm (162) found that feedback about the techniques' effectiveness had no effect on whether subjects deceptiveness or nondeceptiveness could be detected. The theoretical support provided by research on the bogus pipeline indicates that subjects' belief in the technique may be important, and that additional research is needed to assess the effects of belief in the machine on actual polygraph tests.

RESEARCH IMPLICATIONS OF FACTORS AFFECTING VALIDITY

If further research on polygraph testing is carried out, a number of research priorities can be identified on the basis of the present analysis. These priorities include research on the theory of polygraph testing, polygraph techniques, countermeasures, comparison with other techniques, and field-based studies.

Theory

Polygraph testing is premised on the belief that lying produces reliable physiological reactions. Testing the efficacy of this assumption is an important research need. Basic research could examine the physiological reactions to different types of lies and under different conditions of arousal.

Scoring

Research is currently being conducted by the U.S. Army on development of computer scoring systems and more reliable measures of physiological arousal. There is some evidence (e.g., 92) that the validity of polygraph examination decisions is improved if the clinical judgment of examiners is removed (see also, 27) and related evidence that numerical scoring is more accurate (91,133) than nonnumerical scoring. Research should proceed on developing analogs to digital scoring systems. Such research, however, would not address the impact of examiner-examinee interaction, and this area also needs further study.

Question Techniques

Another research priority is to develop additional protocols for question development. Perhaps the most important research need in this regard is to develop and field-test the concealed information test. Basic research and theory (see, e.g., 27,108,136) suggests that such examinations have the highest likelihood of detecting deception, although no field research has yet been conducted to examine their use. Such research should both establish means of constructing GKTs and their validity in actual use.

An additional priority is to develop and test question techniques that may be employed in screening situations. Some examiners for example claim to use a version of CQT for screening examinations (see ch. 2). This application of CQT has not been subjected to scientific tests of validity. In addition, efforts should be devoted to testing the use of CQT with different subject groups and in national security investigations.

Countermeasures

If polygraph testing is to be more widely employed in national security investigations, there is an urgent need for research on countermeasures. Particular priorities would be research on drugs, biofeedback training, and subject gullibility, and motivation. Such research needs to be carried out both in field situations and in the laboratory.

There are a number of drugs that are suspected of lowering ANS arousal and that theoretically may be able to invalidate the results of a polygraph examination or compel an "inconclusive" finding. A first priority is to extend Waid, et al.'s (197), research on meprobamate (which reduced detectability) to other psychoactive drugs.

Biofeedback training, as well as other forms of training have not been investigated, yet their effects on polygraph examinations may be substantial. Subjects' beliefs about the accuracy of the polygraph may also be critical. As suggested by the research on the "bogus pipeline," individuals who believe their underlying thoughts are detectable are more likely to provide truthful responses. The reverse phenomenon seems feasible and it would seem possible to train individuals to believe that the polygraph is ineffective. Such training might be accomplished by providing individuals with false feedback on the polygraph as well as by specific instructions during simulated polygraph examinations. Similarly, subjects who can be easily trained to beat the polygraph may be more desirable as intelligence agents.

Comparison With Other Techniques

Only one study in the available literature (207) systematically compares the polygraph with other investigatory tools. There is a need to examine whether the polygraph provides independent or corroborative evidence and whether the judgments made by polygraph examiners are merely a function of their clinical judgment as investigators, or whether it is a function of the polygraph examination itself.

A complication with this research is that the polygraph functions, in many situations, as a threat. Individuals' fear of taking the examination, in many instances, may lead them to confess or provide incriminating evidence. The threat potential, however, is in part a function of theirs and others' knowledge of research results. If, for example, it became widely known that the polygraph was "beatable," it is likely that this threat would be reduced and, hence, the validity (and utility) of the polygraph would be reduced.

Field Studies

As described in chapters 3, 4, and 5, there are numerous problems with the available field and analog evidence. Field studies are problematic because they can only only be conducted where an independent criterion of guilt or innocence is available. As such, these studies may represent a select sample of cases (e.g., where guilt is overwhelming) and a select set of examiners. Analog studies have a different set of problems and have not adequately motivated subjects or may not have the appropriate number of cases. What is needed is research which deals with the problems of the available field and analog studies.

One "theoretical" solution to the problem of conducting systematic field studies is to conduct "ABSCAM"-like investigations using bogus unauthorized disclosures (instead of bribes) that would enable investigators to set up situations where they have knowledge of who is guilty or innocent. The polygraph could be used to select guilty from innocent with a known base rate and ground truth. Such methods, of course, raise a number of ethical, legal, and pragmatic questions, and it is not clear whether they could provide definitive answers. They could not be used frequently nor with a wide range of techniques/situations. Conducting polygraph research presents serious conceptual and methodological problems; in the absence of such research, however, it will not be possible to develop fully an assessment of the validity of polygraph examinations.

CONCLUSIONS

The description in this chapter of factors affecting validity and potential countermeasures suggests that there is a great deal more to understand about polygraph tests before one can be assured of their validity. Despite our lack of full understanding, however, several factors that affect validity are known. In part, the history of polygraph development over the past 15 to 20 years has been to systematize and improve polygraph testing procedures based on these factors. One central problem, not adequately addressed by either the literature on improvements in validity or countermeasures, is the extent to which these factors affect false negative and positive error rates or affect numbers of inconclusives. For policy purposes, clearly such distinctions and a sense of the magnitude of false decisions is needed. Substantial research, beyond what is currently available, would have to be conducted in order to answer such questions.

Chapter 7

Conclusions

Chapter 7

Conclusions

INTRODUCTION

The primary purpose of this technical memorandum is to evaluate the scientific evidence on the validity of polygraph tests. The memorandum responds to concerns of the Committee on Government Operations, U.S. House of Representatives, about significant changes in Federal Government policy concerning polygraph testing. As discussed in chapters 1 and 3, National Security Decision Directive 84 (NSDD-84), issued by the President on March 11, 1983, authorized executive agencies and departments to require employees to take a polygraph examination in the course of investigations of unauthorized disclosures of classified information. On October 19, 1983, the Department of Justice announced that administration policy would also permit Government-wide polygraph use in preemployment, preclearance, periodic, and aperiodic personnel security screening of employees with access to highly classified information. Draft proposed revisions to Department of Defense (DOD) polygraph regulations (DOD 5210.48) would also authorize the expanded use of polygraph testing as part of personnel security screening of employees with highly sensitive access.

The combined effect of these changes is to authorize substantially expanded use of polygraph examinations by the Federal Government for investigations of specific incidents (i.e., unauthorized disclosures), and, most significantly, for personnel security screening. In addition, NSDD-84, administration policy, and the DOD proposals authorize adverse consequences for refusal to take a polygraph examination.

By letter of February 3, 1983, the Committee on Government Operations asked OTA to assess the scientific evidence on the validity of polygraph testing, based primarily on a critical review and evaluation of existing research. In order to conduct this assessment, OTA studied the actual polygraph examination process, reviewed the results of prior research reviews, analyzed a wide range of relevant field and analog studies, and surveyed Federal agencies as to their polygraph use and any past, present, or planned polygraph research. This chapter highlights the overall scientific conclusions of the OTA evaluation and then discusses in some detail specific scientific conclusions and the implications for recent and proposed changes in Federal policy on polygraph testing.

OVERALL SCIENTIFIC CONCLUSIONS

OTA concluded that, as shown in chapter 2, polygraph testing is, in reality, a very complex process that varies widely in application. Although the polygraph instrument itself is essentially the same for all applications, the purpose of the examination, type of individual tested, examiner training, setting of the examination, and type of questions asked, among other factors, can differ substantially. The instrument cannot itself detect deception. Therefore, polygraph tests require the examiner to develop questions to be asked in each case, compare the physiological response (as measured by the instrument) to the different questions, and infer deception or truthfulness based on these comparisons.

One general type of polygraph question technique (called the control question technique) is commonly used for investigations of specific criminal incidents and has received most of the research attention. Another technique (known as relevant/irrelevant) typically used for preemployment screening and periodic screening purposes has been only minimally researched. Based on a detailed review of these and other question techniques in chapter 2, OTA concluded that there are significant differences, and that the results of research on one technique cannot be generalized to other techniques. Also, differences between techniques are so significant that the results of research on one technique in one application cannot necessarily be extrapolated to other applications. Chapter 2 also reviewed the Federal Government's use of polygraph testing and found that, with the exception of the National Security Agency (NSA) and Central Intelligence Agency (CIA), most current use, even in DOD, is for investigation of specific crimes using the control question technique.

In chapter 3, OTA reviewed the legal, governmental, and scientific controversies over polygraph testing. OTA found that previous debates at the Federal level have focused heavily on whether polygraph testing is scientifically valid. The conclusion of previous congressional inquiries has been that there is little or no scientific basis for the use of polygraph testing. Prior scientific reviews, on the other hand, have contradicted each other, some concluding that polygraph testing is almost 100 percent accurate, others that it is little better than chance. OTA determined that part of the problem in reaching conclusions about polygraph testing validity is that several scientific criteria must be taken into account when assessing validity. Also, previous scientific reviews have not been conducted systematically. In addition, previous reviews, whether legal, governmental, or scientific, have not differentiated polygraph use by type of question technique or application.

OTA conducted its own systematic review of prior research studies on the validity of polygraph testing (see ch. 4 for discussion of field studies of actual polygraph examinations and ch. 5 for discussion of analog or simulation studies). OTA found that there are almost no studies relevant to proposed Federal Government expansion of polygraph testing for preemployment, periodic, or aperiodic screening. This finding has major policy implications discussed later. OTA also found that, even among the rather extensive studies of the control question technique in criminal investigations, there is a wide range of accuracy (and thus, inconclusive and error) rates. OTA concluded that this accuracy range could be partially explained by variations in research design but perhaps to a greater extent, as is discussed in chapter 6, by differences in examiners, examinees, question techniques, and conditions of testing.

OTA concluded, therefore, that no overall measure or single statistic of polygraph validity can be established based on available scientific evidence. The amount and quality of the evidence depends on the design and conduct of specific studies and the particular application researched. Some applications (e.g., the use of the polygraph in criminal investigations) have been fairly heavily researched, while others (e.g., polygraph use in preemployment screening) have had very little research attention.

Further, regardless of whether polygraph testing is used in specific-incident investigations or personnel screening, OTA concluded that polygraph accuracy may also be affected by a number of factors: examiner training, orientation, and experience; examinee characteristics such as emotional stability and intelligence; and, in particular, the use of countermeasures and the willingness of the examinee to be tested. In addition, the basic theory (or theories) of how the polygraph test actually works has been only minimally developed and researched.

In sum, OTA concluded that there is at present only limited scientific evidence for establishing the validity of polygraph testing. Even where the evidence seems to indicate that polygraph testing detects deceptive subjects better than chance (when using the control question technique in specific-incident criminal investigations), significant error rates are possible, and examiner and examinee differences and the use of countermeasures may further affect validity.

More specific scientific conclusions and the implications for recent and proposed changes in Federal policy on polygraph testing are presented below. The discussion is organized in terms of conclusions and implications, first, for specific-incident investigations and personnel security screening use of the polygraph; second, for polygraph countermeasures and for the voluntary nature of testing; and finally, for further research.

SPECIFIC SCIENTIFIC CONCLUSIONS IN POLICY CONTEXT

Specific-Incident Criminal Investigations

A principal use of the polygraph test is as part of an investigation (usually conducted by law enforcement or private security officers) of a specific situation in which a criminal act has been alleged to have, or in fact has, taken place. This type of case is characterized by a prior investigation that both narrows the suspect list down to a very small number, and that develops significant information about the crime itself. When the polygraph is used in this context, the application is known as a specific-issue or specific-incident criminal investigation.

Results of OTA Review

The application of the polygraph to specific-incident criminal investigations is the only one to be extensively researched. OTA identified 6 prior reviews of such research (summarized in ch. 3), as well as 10 field and 14 analog studies that met minimum scientific standards and were conducted using the control question technique (the most common technique used in criminal investigations; see chs. 2, 3, and 4). Still, even though meeting minimal scientific standards, many of these research studies had various methodological problems that reduce the extent to which results can be generalized. The cases and examiners were often sampled selectively rather than randomly. For field studies, the criteria for actual guilt or innocence varied and in some studies were inadequate. In addition, only some versions of the control question technique have been researched, and the effect of different types of examiners, subjects, settings, and countermeasures has not been systematically explored.

Nonetheless, this research is the best available source of evidence on which to evaluate the scientific validity of the polygraph for specific-incident criminal investigations. The results (for research on the control question technique in specific-incident criminal investigations) are summarized below:

- Six prior reviews of field studies:
  - —average accuracy ranged from 64 to 98 percent.
- Ten individual field studies:
  - —correct guilty detections ranged from 70.6 to 98.6 percent and averaged 86.3 percent;
  - —correct innocent detections ranged from 12.5 to 94.1 percent and averaged 76 percent;
  - —false positive rate (innocent persons found deceptive) ranged from 0 to 75 percent and averaged 19.1 percent; and
  - —false negative rate (guilty persons found nondeceptive) ranged from 0 to 29.4 percent and averaged 10.2 percent.
- Fourteen individual analog studies:
  - —correct guilty detections ranged from 35.4 to 100 percent and averaged 63.7 percent;
  - —correct innocent detections ranged from 32 to 91 percent and averaged 57.9 percent;
  - —false positives ranged from 2 to 50.7 percent and averaged 14.1 percent; and
  - —false negatives ranged from 0 to 28.7 percent and averaged 10.4 percent.

The wide variability of results from both prior research reviews and OTA's own review of individual studies makes it impossible to determine a specific overall quantitative measure of polygraph validity. The preponderance of research evidence does indicate that, when the control question technique is used in specific-incident criminal investigations, the polygraph detects deception at a rate better than chance, but with error rates that could be considered significant.

The figures presented above are strictly ranges or averages for groups of research studies. Another selection of studies would yield different results, although OTA's selection represents the set of studies that met minimum scientific criteria. Also, some researchers exclude inconclusive results in calculating accuracy rates. OTA elected to include the inconclusives on the grounds that an inconclusive is an error in the sense that a guilty or innocent person has not been correctly identified. Exclusion of inconclusives would raise the overall accuracy rates calculated. In practice, inconclusive results may be followed by a re-test or other investigations.

Relevance to NSDD-84 and Administration Policy

While the results of the OTA review indicate that the control question technique has some validity in criminal investigations, there is only a limited scientific basis for generalizing the results of the OTA review to the context of NSDD-84 and the October 19, 1983, administration policy on polygraph use. NSDD-84 and administration policy authorize the use of the polygraph in administrative as well as criminal investigations of unauthorized disclosures of classified information.

First, there is no validity research directly on the use of the polygraph in unauthorized disclosure investigations. The subject matter and perhaps subjects of these investigations will vary from the typical criminal investigation as might the conditions and techniques of testing and use of countermeasures.

Second, the investigative conditions authorized by NSDD-84 and administration policy may be quite different from conditions under which prior research was conducted. NSDD-84 does not specify what type of investigative procedures will be followed, how subjects will be selected or identified, who will conduct the examinations, or what question techniques will be used. Administration policy provides some specific guidelines such as requiring that polygraph testing be used only when "other information or means of investigation have produced a substantial objective basis for seeking to examine the employee" and there is "no other reasonable means of resolving the matter" (185a). However, in general, the extent to which employees will be requested or required to take polygraph examinations in unauthorized disclosure investigations is largely left to the discretion of agency heads.

Third, even the Federal Bureau of Investigation (FBI) has concluded that, "to date, no methodologically adequate study of control question techniques has been reported. . . . Inferences regarding the validity of control question examinations . . . rest upon the results of laboratory studies conducted under highly dissimilar conditions." The FBI is planning its own validity research.

On the other hand, to the extent polygraph use in unauthorized disclosure investigations is similar to the way the polygraph is used in criminal investigations, there is at least some although far from conclusive scientific basis for polygraph validity.

Large-Scale Screening

The polygraph test is used by some private firms and on rare occasions by some Federal agencies to screen a large number of people in connection with the investigation of a crime. Unlike the typical specific-incident criminal investigation, in a large-scale screening investigation, typically the suspect list has not been narrowed down to one or a few persons and only limited information about the crime is available.

NSDD-84 appears to permit such use of the polygraph in unauthorized disclosure investigations, although the actual extent of NSDD-84 is unclear. Administration policy appears to be ambivalent. While on the one hand providing guidelines for "carefully limited use of the polygraph," the policy implies that DOD polygraph regulations are acceptable. DOD regulations have been used, albeit infrequently, to authorize polygraph screening of large numbers of individuals (ranging from about 2 dozen up to 80) in investigation of specific incidents.

There is no scientific basis for generalizing the results of the OTA review to establish polygraph validity in this large-scale screening application. *First,* no scientifically acceptable research has been conducted on large-scale specific-incident screening use of the polygraph. *Second,* the screening conditions here are likely to vary even more from the conditions of the research studies reviewed by OTA. For one thing, much less information is likely to be known about circumstances surrounding an unauthorized disclosure and possible suspects. This could translate into differences in the questions used, the behavior of the polygraph examiner, the motivation and response of the subject, and the effectiveness of countermeasures.

Third, the large-scale screening use of polygraph testing theoretically can be expected to result in significantly higher error rates than when the list of suspects is narrowed down to a very small number, as in a typical criminal investigation. The screening use of polygraph tests is most dependent on the so-called base rate of guilt, i.e., the percentage of the group of persons being screened that has engaged in the criminal (or otherwise proscribed) activity. If the percentage of guilty is small, say 5 percent (1 guilty person out of every 20 persons screened, or 50 out of 1,000), then even assuming a very high (95 percent) polygraph validity rate, the predictive value of the screening use of the polygraph would only be 50 percent. That is, for each 1,000 individuals screened, about 47 out of the 50 guilty persons would be correctly identified as deceptive, but 47 out of the 950 innocent persons would be incorrectly identified as deceptive (false positives). Thus of the 94 persons identified as deceptive, one-half would be innocent persons. For every person correctly identified as deceptive, another person would be incorrectly identified.

As another example, if a lower polygraph validity rate is assumed (say 90 percent), then the predictive value would be expected to drop to about 33 percent. That is, for every person correctly identified as deceptive, two persons would be incorrectly identified (false positives).

These are, of course, hypothetical examples, and have not been systematically investigated in either field or analog research, although some reviewers (e.g., Ben-Shakhar (28)) have carefully worked through a number of possibilities. Also, operating procedures of Federal agencies (e.g, quality control review, consideration of other investigatory information) might catch, correct, or minimize erroneous polygraph decisions.

Nonetheless, the FBI, which outside of DOD and CIA, is the principal Federal agency that conducts polygraph examinations, believes that large-scale screening is not an appropriate use of polygraph testing. FBI regulations prohibit the "use of the polygraph for dragnet-type screening of large numbers of suspects or as a substitute for logical investigation by conventional means" [FBI Polygraph Regulation 13-22.2 (2), 1980].

Personnel Security Screening

Draft revisions to the DOD polygraph regulations would authorize the use of polygraph tests to determine initial and continuing eligibility of DOD civilian, military, and contractor personnel for access to highly classified information (Sensitive Compartmented Information and/or special access). The use of polygraph tests to determine continuing eligibility would be on an aperiodic (i.e., irregular) basis (181). These are all known as personnel security applications of the polygraph. In addition, administration policy announced on October 19, 1983, would permit Government-wide use of polygraph tests in personnel security screening of employees (and applicants for positions) with access to highly classified information. The new policy provides agency heads with the authority to give polygraph examinations on a periodic or aperiodic basis to employees with highly sensitive access.

Results of OTA Review

Personnel security screening involves a different type of polygraph test than specific-incident investigations, and very little screening research has been conducted. Three studies were cited by the intelligence agencies (NSA and CIA) as providing support for personnel security use of polygraph tests.

A 1975 field study (6) of polygraph screening of government job applicants (from an unidentified Federal agency) showed high consistency in readings of physiological arousal by different examiners. But this study concluded nothing about validity.

In a 1981 analog study (43) of preemployment screening use, 75 percent of the responses of deceptive individuals were detected accurately. Twenty-five percent were detected incorrectly. Any conclusions based on this study must be limited by the fact that the subjects were students, the questions and context had nothing to do with national security, and the test format was atypical of personnel screening examinations.

A 1980 survey conducted by the Director of the Central Intelligence Security Committee concluded that the polygraph was the most productive of all background investigation techniques. However, this was a utility study not a validity study, and had many limitations and qualifications. For example, the criteria for case selection were not stated and there was no independent verification of the cases that were resolved. Also, the polygraph was used only after a thorough investigation based on other sources had taken place (see ch. 4 for further discussion).

OTA inquiries to all DOD components using the polygraph identified only one DOD research study on personnel screening use of the polygraph (16). The results of this study raise more questions than they answer, and certainly do not provide support for high polygraph validity in a screening situation. The limitations of the study reduce its applicability, but it is the only DOD polygraph screening research known to OTA. OTA inquiries to other executive agencies and departments using the polygraph identified no research on personnel security screening use of the polygraph.

OTA recognizes that the administration as well as NSA, CIA, and DOD believe that the polygraph is a useful screening tool. However, OTA concluded that the available research evidence does not establish the scientific validity of the polygraph for this purpose.

In comments to OTA, CIA agreed that the cumulative unclassified research evidence reviewed by OTA is not directly relevant to national security applications. However, CIA does claim to have classified research to support their use of polygraph tests. OTA did not review this research. No other Federal agency, including NSA, has claimed to have relevant research results that were not available for OTA review on an unclassified basis.

False Positives

One area of special concern in personnel security screening is the incorrect identification of innocent persons as deceptive. All other factors being equal, the low base rates of guilt in screening situations would lead to high false positive rates, even assuming very high polygraph validity. For example, a typical polygraph screening situation might involve a base rate of one guilty person (e.g., one person engaging in unauthorized disclosure) out of 1,000 employees. Assuming that the polygraph is 95 percent valid, then, the one guilty person would be identified as deceptive but so would 50 innocent persons. The predictive validity would be about 2 percent. Even if 99 percent polygraph validity is assumed, there would still be 10 false positives for every correct detection of a guilty person.

Again, these are hypothetical examples that have not been systematically studied in field or analog research. NSA claims that they in fact have experienced a very low false positive rate and that, in any event, polygraph test results are only one factor in making decisions and are subject to quality control checks and other reviews. It appears that NSA (and possibly CIA) use the polygraph not to determine deception or truthfulness per se, but as a technique of interrogation to encourage admissions. NSA has stated that the agency "does not use the 'truth v. deceptive' concept of polygraph examinations commonly used in criminal cases. Rather, the polygraph examination results that are most important to NSA security adjudicators are the data provided by the individual during the pretest or posttest phase of the examination" (187).

The validity of the polygraph as used by NSA has not been researched. And, in general, this kind of application is potentially different in so many ways from the polygraph use in specific-incident criminal investigations (e.g., with respect to type of questions asked and question techniques employed) that results of the OTA research review previously discussed cannot be generalized to the NSA situation.

False Negatives/Countermeasures

The primary purpose of polygraph testing under NSDD-84, the DOD revised regulations, and administration policy is to detect persons who have or intend to participate in proscribed activities (e.g., unauthorized contact with a foreign agent, disclosure of classified information). A concern with false negatives (guilty persons incorrectly identified as nondeceptive) is that, apart from any errors inherent in the polygraph test itself, the guilty person may be able to escape detection through the use of countermeasures.

Theoretically, polygraph testing—whether for personnel security screening or specific-incident investigations—is open to a large number of countermeasures, including physical movement or pressure, drugs, hypnosis, biofeedback, and prior experience in passing an exam. The research on polygraph countermeasures has been limited and the results—while conflicting—suggest that validity may be affected. Further, some research (e.g., 75) suggests that polygraph examiners may not be able to easily detect certain physical countermeasures. The research results for drug and psychological countermeasures are mixed. The possible effects of countermeasures are particularly significant to the extent that the polygraph is used and relied on for national security purposes, since even a small false negative rate could have serious consequences. In addition, those individuals who the Federal Government would most want to detect (e.g., for national security violations) may well be the most motivated and perhaps the best trained to avoid detection.

Voluntary v. Involuntary

As currently used in the Federal Government, with few exceptions, polygraph examinations are voluntary. That is, a person cannot be forced to take a polygraph test against his or her will. A refusal to take a polygraph test does not, or at least is not supposed to, result in adverse consequences. The only exceptions are NSA (and by extension, CIA) and, under certain conditions, the FBI. NSA notes that "the polygraph examination is part of the Agency's security processing. Failure to complete processing may result in failure to be accepted for employment" (187). FBI regulations require that "polygraph examinations will be administered only to individuals who agree or volunteer to take an examination" [FBI Regulation 13-22.2(3)]. The only exception is for certain FBI employees and applicants under specified circumstances where "a refusal to be examined by polygraph may lead to an adverse inference being drawn."

The DOD proposal would provide that refusal to take a polygraph examination, when established as a requirement for selection or assignment or as a condition of access, may result in adverse consequences for the individual. These include nonselection for assignment or employment, denial or revocation of clearance, or reassignment to a nonsensitive position. NSDD-84 also provides that refusal to take a polygraph test may result in adverse consequences such as administrative sanctions and denial of security clearance. And administration policy authorizes denial of clearance, transfer or reassignment, and, under some circumstances, termination of employment for refusal to take a polygraph test.

Under these conditions, polygraph examinations would not be voluntary in the strict sense, since a refusal could result in penalties. Apart from the ethical and perhaps legal implications, which OTA did not address, conducting polygraph tests on this basis could affect test validity. It is generally recognized that, for the polygraph test to be accurate, the voluntary cooperation of the individual is important. For example, NSA has stated that, in conducting screening examinations, "[t]he full cooperation of the individual taking the test is essential or the results will be inconclusive." The polygraph only detects physiological arousal, and under involuntary conditions, the arousal response of the examinee may be very difficult or impossible to interpret. However, no direct research on this topic was identified. Overall, OTA concluded that imposing penalties for not taking a test may create a de facto involuntary condition that increases the chances of invalid or inconclusive test results.

Further Research

OTA concluded that, to the extent that polygraph testing is going to continue to be used by the Federal Government, further research is needed. Possible research priorities include the following.

Polygraph Theory

The basic theory of polygraph testing is only partially developed and researched. The most commonly accepted theory at present is that, when the person being examined fears detection, that fear produces a measurable physiological reaction when the person responds deceptively. Thus, in this theory, the polygraph instrument is measuring the fear of detection rather than deception per se. And the examiner infers deception when the physiological response to questions about the crime or unauthorized activity is greater than the response to other questions. However, this theory has been challenged by some psychologists and others who believe that various factors—e.g., the examinee's intelligence level, psychological health, emotional stability, and belief in the "machine"—may, at least theoretically, affect the physiological response.

OTA concluded that a stronger theoretical base is needed for the entire range of polygraph applications, including current and proposed Federal Government applications. Basic polygraph research should consider the latest research from the fields of psychology, physiology, psychiatry, neuroscience, and medicine; comparison among question techniques; and measures of physiological response.

Criminal Investigation Validity

There are still many unanswered questions about the validity of use of the polygraph in specific-incident criminal investigations. A planned FBI-Secret Service validity study is intended to meet this need. However, OTA did not review the research plan, which would benefit from an independent review by the scientific community and others before the research approach is finalized. Such a review would help ensure that the research design is as scientifically sound as possible. Also, the U.S. Army's current 10-year research program to develop a new state-of-the-art polygraph instrument should be reevaluated to determine if research priorities and direction need adjustment. As it stands now, validity issues will not be addressed by the Army research until the late 1980's.

Personnel Security Screening Validity

Given the almost total lack of research on this application, further research is clearly necessary if there is to be any possibility of establishing a scientific basis for the personnel security screening use of polygraph testing.

Research on Polygraph Countermeasures

Since NSA and CIA are already heavily dependent on the polygraph, their use alone justifies an intensified research effort on countermeasures. NSA and the U.S. Army Intelligence and Security Command are planning such research, but the level of effort appears low (e.g., $65,000 pilot study in NSA) considering the consequences of false negatives.

CONCLUDING COMMENT

A major reason why scientific debate over polygraph validity yields conflicting conclusions is that the validity of such a complex procedure is very difficult to assess and may vary widely from one application to another. The accuracy obtained in one situation or research study may not generalize to different situations or to different types of persons being tested. Scientifically acceptable research on polygraph testing is hard to design and conduct.

Advocates of polygraph testing argue that thousands of polygraphs have been conducted which substantiate its usefulness in criminal or screening situations. Claims of usefulness, however, are often dependent on information (e.g., confessions and admissions) obtained before or after the actual test, and on its perceived value as a deterrent.

The focus of the OTA technical memorandum is not whether the polygraph test has been useful, but whether there is a scientific basis for its use. OTA concluded that, while there is some evidence for the validity of polygraph testing as an adjunct to typical criminal investigations of specific incidents, and more limited evidence when such investigations extend to incidents of unauthorized disclosure. However, there is very little research or scientific evidence to establish polygraph test validity in large-scale screening as part of unauthorized disclosure investigations, or in personnel security screening situations, whether they be preemployment, preclearance, periodic or aperiodic, random, or "dragnet." Substantial research beyond what is currently available or planned would have to be conducted in order to fully assess the scientific validity of the NSDD-84, DOD, and administration polygraph proposals.

Appendixes

Appendix A

Informed Consent Forms

POLYGRAPH EXAMINATION WAIVER

Place: ______________________________

Date & Time: ______________________

I, ____, have been requested by Special Agent ____ of the Naval Investigative Service to submit to a polygraph examination relative to my (knowledge of) (participation in)

____

With respect to that request, I have been advised:

(a) that I have the right to consult with a lawyer prior to making any decision concerning the examination; ______

(b) that the polygraph examination will be conducted only with my prior written consent; ________

(c) that no adverse action will be taken against me solely on the grounds that I refuse to consent to this examination; __________

(d) that the area in which the examination is to be conducted (does) (does not) contain a two-way mirror or similar device; ________

(e) that the area in which the examination is to be conducted (does) (does not) contain a camera; __________

(f) that the area in which the examination is to be conducted (does) (does not) contain an electronic audio recording device and the polygraph examination (will) (will not) be monitored. __________

With an understanding of the above conditions, I have decided that I do not desire to consult with a lawyer at this time. I freely consent to be examined by polygraph and I agree to cooperate fully with the examiner during that examination. I make these decisions freely and voluntarily and they are made with no threats having been made or promises extended to me.

____

(Signature)

Time: ______________________________

Witnessed:

______ ______________________

NISFORM 010-F/04-80 GPO 885-471

POLYGRAPH EXAMINATION STATEMENT OF CONSENT	For use of this form, see AR 195-6; the proponent agency is U.S. Army Criminal Investigation Command.	
PLACE	DATE	TIME
STATEMENT OF CONSENT OF *(Name, Grade, and SSN)*	PLACE OF BIRTH DATE OF BIRTH	

(Strike out inapplicable portions indicated in parentheses.)

In the presence of the witness(es) whose signature(s) appear(s) below, Article 31 of the Uniform Code of Military Justice and The Fifth Amendment to the Constitution of the United States have been explained to me by

who informed me that he is a polygraph examiner of the United States Army. He has informed me that this statement is being completed in connection with

He explained to me the nature of the polygraph examination and told me that I cannot be required to take such examination without my consent. He explained to me that I have the right to consult with counsel prior to this examination and to have counsel present to observe the examination. He explained that this counsel may be civilian counsel retained at my expense, or counsel appointed for me at no expense to me (,or if a member of the United States Armed Forces, that I may select military counsel of my choice if such counsel is reasonably available). I (do) (do not) want to consult with counsel prior to this examination. I (do) (do not) want to have counsel present to observe the examination. I was further advised that the examination room (does) (does not) contain a two-way mirror or observation device and that the examination (will) (will not) be monitored or recorded. He explained to me that I do not have to make any statement whatsoever but that any statement I do make may be used as evidence against me in a trial by court-martial (if subject to the Uniform Code of Military Justice) or in any other military or judicial proceedings. Understanding my unqualified right to refuse, I, understand that I will be questioned prior to, during and after the instrument portion(s) of the polygraph examination and do hereby, this date, voluntarily and without duress, coercion, unlawful inducement, or promise of reward, consent to a polygraph examination.

WITNESSES	SIGNATURE OF PERSON TO BE TESTED
SIGNATURE	
TYPED NAME ORGANIZATION AND/OR ADDRESS	ORGANIZATION AND/OR ADDRESS
SIGNATURE	SIGNATURE OF EXAMINER
TYPED NAME ORGANIZATION AND/OR ADDRESS	
EXHIBIT NUMBER	TYPED NAME AND ORGANIZATION

DA FORM 1 AUG 67 2801

REPLACES EDITION OF 1 JAN 64, WHICH IS OBSOLETE.

☆U.S. Government Printing Office: 1980—210-981/9082

Appendix B

Results of the OTA Survey of Federal Government Polygraph Testing

Introduction

In May 1983, OTA surveyed selected Federal Government agencies including the Departments of State, Defense (DOD), Treasury, and Justice, Central Intelligence Agency (CIA), Office of Personnel Management, and U.S. Postal Service (USPS), with respect to their use of polygraph testing. The survey requested detailed information about agencies' current and past use of polygraph testing and research conducted or planned by the agency. The request for information was sent to all Federal agencies believed to conduct polygraph examinations. A follow-up survey was sent, in July 1983, with respect to use of polygraph testing in unauthorized disclosure investigations.

Results of the survey are described below. All agencies responded to OTA's inquiry; however, the CIA considers all such operational and research information to be classified. In addition, the results do not include information from the Customs Service (a Department of the Treasury component), Department of Health and Human Services, and Tennessee Valley Authority, which conduct a limited but unknown number of polygraph examinations. OTA supplemented the survey results with site visits to polygraph facilities at the U.S. Army, National Security Agency (NSA), and Federal Bureau of Investigation (FBI), and discussions with officials from several Federal agency polygraph programs.

Number of Polygraph Examinations

For 1982, the agencies reported conducting a total of 22,597 individual polygraph examinations. Of this total, 18,301 examinations were conducted by DOD component agencies, including the Army, Navy, Air Force, Marines, and NSA. Individual agency totals are shown in table B-1. NSA conducts the largest number of examinations, 43 percent of the total. Next, in terms of number of tests, is the Army Criminal Investigation Command, followed by the Air Force Office of Special Investigations, Naval Investigative Service, and FBI. The NSA and the Air Force have steadily increased the number of examinations conducted each year during the 1980-82 period, while the number of polygraph examinations appears to be relatively stable over this period in other agencies.

However, long-term trends in the number of polygraph examinations show a substantial increase since 1973. In fact, the total number of examinations in 1982 was more than triple the 1973 total (22,597 examinations in 1982 compared to 6,946 in 1973) and actually surpassed the previous known high (19,796 in 1963, excluding NSA). As illustrated below, the FBI, Air Force, and NSA experienced the largest absolute increases in polygraph examinations over the 1973-82 period.

	Number of examinations conducted		
Agency	*Fiscal year 1963*	*Fiscal year 1973*	*Fiscal year 1982*
Army CIC	4,400	2,028	3,731
Army ISC	8,094		279
Navy	1,200	665	1,337
Air Force	1,912	482	3,019
Marines	812	62	263
NSA	Not available	3,081	9,672
Other DOD	140	6	0
DOD subtotals	16,558	6,325	18,301
FBI	2,314	79	2,463
DEA			211
SS	65	50	714
BATF			256
USPS	338	485	652
Other	521	7	0
Totals	19,796	6,946	22,597

SOURCE: Data from the Office of Technology Assessment, 1982; 1973 and 1963 data from U.S. Congress, House of Representatives, Committee on Government Operations, reports, *The Use of Polygraphs and Similar Devices by Federal Agencies,* 1976 and *Use of Polygraphs as Lie Detectors by the Federal Government,* 1965.

Number of Polygraph Examiners

For 1982, agencies reported employing a total of 209 polygraph examiners. Of these examiners, the majority (130) were employed by DOD component agencies. Individual agency totals are shown in table B-1. The U.S. Army has the largest number of examiners, followed closely by the FBI, and then by the U.S. Air Force and NSA. The reason that the number of examiners is not directly related to the number of examinations is that examinations are conducted by agencies for different purposes and under different conditions. For example, NSA examinations are conducted for screening purposes in a central location; in contrast, Army examinations are conducted primarily as part of criminal investigations, and examiners frequently travel to sites within a geographic region.

Table B-1.—Number of Polygraph Exams and Examiners

	Number of polygraph exams			Number of examiners		
	Fiscal year			Fiscal year		
Agency	1980	1981	1982	1980	1981	1982
Department of Defense:						
Army Criminal Investigation Command	3,977	3,832	3,731	39	42	44
Army Intelligence and Security Command	230	260	279	9	9	12
Naval Investigative Service	1,317[a]	1,185[a]	1,337[a]	11[a]	12[a]	14[a]
Air Force Office of Special Investigations	1,474[a]	1,485[a]	3,019[a]	26[a]	29[a]	28[a]
Marines	376[a]	245[a]	263[a]	8[a]	6[a]	6[a]
National Security Agency	5,676[a]	7,418[a]	9,672[a]	13[a]	30[a]	26[a]
Subtotals	13,050	14,425	18,301	106	128	130
Department of State	Does not conduct polygraph exams					
Department of Justice:						
Federal Bureau of Investigation	2,121	2,162	2,463	NA	NA	40
Drug Enforcement Administration	230	200	211	NA	NA	14
Department of Treasury:[b]						
Secret Service	NA	466	714	NA	NA	16
Bureau of Alcohol, Tobacco and Firearms	176	254	256	4	4	4
U.S. Postal Service	714	725	652	NA	NA	5
Central Intelligence Agency	Does conduct polygraph exams but specific operational information is classified					
Office of Personnel Management	Does not conduct or use polygraph exams					
Totals	16,291	18,232	22,597			209

[a] Calendar year.
[b] Excludes Customs Service.
NA = Not available

Other Federal Agency Polygraph Users

The Federal agencies listed in table B-1 are the primary users of the results of polygraph tests conducted by their personnel. However, these agencies reported that during 1980-82, polygraph examinations were also conducted by their staff for other Federal agencies, both those with polygraph capability and those without. A listing of the number of examinations conducted for agencies that do not employ their own polygraph staffs follows:

Exams conducted by	*Exams conducted for*	*Number of exams 1980-82*
Army, CIC	Department of State	26
	Internal Revenue Service	1
	Defense Investigative Service	1
	Department of Defense (other)	14
Army, ISC	Defense Intelligence Agency	7
Navy	Coast Guard	1
	General Services Administration	1
	Department of State	2
Air Force	Defense Investigative Service	16
	Defense Intelligence Agency	21
	Coast Guard	1
	Department of State	1
Marines	None	
NSA	DOD components	Data not available
FBI	Bureau of Prisons	39 (1982)
	Other Agencies	10 per year
DEA	Immigration and Naturalization Service	2 (1981-1982)
	U.S. Marshall's Office	3 (1981-1982)
	Department of State	2 (1981-1982)
	Internal Revenue Service	1 (1981-1982)
Secret Service	Internal Revenue Service	Specific data not available, but total is less than 8 percent of all Secret Service exams.
	U.S. Attorney's Office	
	Department of Treasury	
	Department of Agriculture	
	Federal Reserve Bank	
BATF	Other Agencies (very limited)	Data not available
USPS	Internal Revenue Service	4
	U.S. Marshall's Office	1
	U.S. Congress	1

The polygraph use by these other agencies represents a very small percentage of total Federal agency use.

Purpose of Polygraph Examinations

As shown in table B-2, with the exception of NSA, over two-thirds of Federal agency use of the polygraph is for criminal investigative purposes. In the major Federal polygraph user agencies, such as the Army, Navy, Air Force, and FBI, over 90 percent of polygraph use is for criminal investigations, for example in the verification of information provided by suspects, victims, and witnesses. The one exception, for which data are available, is NSA. About two-thirds of NSA poly-

Table B-2.—Purpose of Polygraph Exam

		Criminal investigation	Counter intelligence	Intelligence	Other		
Department of Defense:							
Army Criminal Investigation Command	1980	3,968	—	—	9 polygraph examiner applicants		
	1981	3,820	—	—	12 polygraph examiner applicants		
	1982	3,713	—	—	19 polygraph examiner applicants		
					Personnel security	Limited access	Polygraph applicants
Army Intelligence and Security Command	1980	NA	44	NA	0	58	5
	1981	NA	33	NA	9	34	1
	1982	NA	78	NA	58	62	2
Navy	1980[a]	1,209	30	78			
	1981[a]	1,049	50	86			
	1982[a]	1,210	45	82			
Air Force	1980[a]	1,296	NA	NA			
	1981[a]	1,298	NA	NA			
	1982[a]	1,750	NA	NA			
Marines		NA	—	—	Polygraph examiner applicants		
National Security Agency		NA	NA	NA	Applicant screening		
Department of Justice:							
Federal Bureau of Investigation	1980-82	6,038	474	—	234 personnel security		
Drug Enforcement Administration	1980-82	449	—	—	192 internal investigations		
Department of the Treasury:							
Secret Service	1982	562	—	65	59 other agency 16 bond 12 inspection		
Bureau of Alcohol, Tobacco and Firearms	1980-82	686	—	—			
U.S. Postal Service	1980-82	2,091	—	—			

[a] Calendar Year.
NA = Not available.

graph examinations are for applicant screening; i.e., for use in personnel security evaluations of applicants for employment. In 1982, OTA estimates that NSA conducted about 6,700 applicant screening polygraph exams. No other Federal agency, except CIA, conducts routine applicant screening polygraph exams. CIA, as noted above, did not provide information on the purpose of their exams. However, public information available from a report of the Permanent Select Committee on Intelligence, U.S. House of Representatives (173), indicates that the CIA utilizes polygraph tests as part of its applicant screening.

The following agencies also conduct a small number of polygraph exams for counterintelligence and/or intelligence purposes (see table B-2 for estimates): Army Security and Intelligence Command, Navy, Air Force, NSA, FBI, and Secret Service. Other miscellaneous purposes for polygraph exams are listed in table B-2.

Use of Polygraph in Unauthorized Disclosure Cases

Polygraph exams are used by several Federal agencies in connection with the investigation of the unauthorized disclosure of sensitive or classified information, however, such use at present is limited.

Federal agencies responding reported the following polygraph use in unauthorized disclosure cases over the 1980-82 period:

Agency	*Number of polygraph examinations (1980-82)*
Army, CIC	Very few
Army, ISC	1
Navy	78
Air Force	112
Marines	0
NSA	Data not available
State Department	0
FBI	26 (since 1978)
DEA	33
Secret Service	11
BATF	0
USPS	0

For agencies providing detailed statistics, the results of the exams were as follows:

	Deceptive	Not deceptive	Inconclusive	No opinion	Deceptive confirmed
Army, ISC	0	1	0	0	0
Navy	26	51	1	0	18
Air Force	26	85	1	0	21
FBI	16	10	0	0	14
DEA	2	31	0	0	Data not available
Secret Service	0	11	0	0	0

Confirmation of deceptive exam results was primarily though a pre- or post-test confession or admission. Very few of the not deceptive test results were confirmed. Except for the FBI, information was not available on what action, if any (e.g., administrative sanction, removal of security clearance, criminal prosecution), was taken based on the deceptive exam results. The FBI reports that in 12 closed cases, deceptive examination results contributed (at least in part) to 3 convictions, 1 dismissal, 1 disciplinary action, 3 resignations, 3 censures, and 1 voluntary retirement.

Polygraph Examiner Training and Techniques

Federal agencies reported a high degree of consistency in the training of and techniques used by Federal polygraph examiners. All agencies, except NSA, reported that examiners are required to be graduates of the 12-week U.S. Army Polygraph Training Course at Ft. McClellan, Ala. (a component of the U.S. Army Military Police School). NSA requires examiners to be graduates of either the U.S. Army School or the Maryland Institute of Criminal Justice. All examiners are required to have at least 2 years investigative experience. USPS requires 3 years investigative experience, the Secret Service requires 4 years investigative experience, and the Navy, FBI, and BATF require 5 years. In addition, all examiners are required to have an undergraduate degree from an accredited college. The Drug Enforcement Agency (DEA), Secret Service, and BATF require examiners to participate in an advanced or refresher course every year; DOD components and the FBI require such participation every 2 years; and USPS requires such participation every 3 years. All examiners are required to complete an internship or probationary period after graduation from polygraph school.

With respect to examiner technique, examiners at all agencies reporting except NSA make primary use of one or more control question techniques. The modified general question and zone of comparison are the most frequently used control question techniques. Examiners at most agencies also use the peak of tension technique (a concealed information technique). At NSA, the relevant/irrelevant technique is the most frequently used. The Army Intelligence Command, FBI, DEA, Secret Service, and BATF also use the relevant/irrelevant technique to a limited extent. All agencies reported that examiners use a standardized numerical scoring system for interpreting results of exams conducted with a control question technique. For exams conducted with the relevant/irrelevant technique, the examiner looks for significant, consistent reactions. See chapter 2 for further discussion of question techniques.

Methods of Quality Control

All Federal agencies reported that essentially the same polygraph instruments and physiological measures are employed in conducting polygraph examinations. All Federal agencies use primarily Stoelting and Lafayette polygraph instruments (purchased from private manufacturers). The physiological measures include respiration (breathing), perspiration (galvanic skin response), and cardiovascular (blood pressure and pulse rate).

Agencies also indicated that all polygrams (charts) are reviewed independently by a supervisor and/or a polygraph coordinator at a headquarters location. This quality control review includes checking the original examiner's chart interpretation as well as reviewing question construction and other aspects of the exam. Agencies vary in the specifics of their quality control process, but any disagreement between the chart interpretations of the original examiner and quality control examiner usually requires a reexamination.

Length of Polygraph Examinations

Agencies reported that the length of polygraph examinations ranges from about 1.5 to 4 hours, as indicated in table B-3.

Results of Examinations and Subsequent Confirmation

The results of polygraph examinations vary widely among Federal agencies. The number of deceptive examination results ranges from about 10 percent of total exams (for USPS in 1981) to about 69 percent (Army Criminal Investigation Command, 1980), with most agencies in the 40 to 60 percent deceptive range. See table B-3 for agency specific statistics.

Confirmation of results also varies widely, as shown in table B-4. Independent confirmation rates for deceptive exam results range from about 25 percent for the Marines to 70 to 80 percent for the Army Criminal In-

Table B-3.—Length and Results of Exams

	Average length of exam			Results of exams											
	Fiscal year 1980	Fiscal year 1981	Fiscal year 1982	Fiscal year 1980 percent				Fiscal year 1981 percent				Fiscal year 1982 percent			
				D	ND	I	NO	D	ND	I	NO	D	ND	I	NO
Department of Defense:															
Army Criminal Investigation Command	2:53	2:54	3:03	68.7	21.4	0.6	9.2	66.3	24.1	0.9	8.7	64.0	28.3	0.6	7.1
Army Intelligence and Security Command	4 hours average			34	59	4	3	33	62	2	3	34	59	3	4
Air Force				50	50	—	—	50	50	—	—	50	50	—	—
Marines	At least 1.5 hours			51	NA	NA	NA	51	NA	NA	NA	52	NA	NA	NA
National Security Agency	1.5 to 2 hours average			NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Department of Justice:															
Federal Bureau of Investigation	NA			3,527 deceptive out of 6,646 total exams for FY 80-82											
Drug Enforcement Administration	2 to 3 hours average			171 deceptive out of 641 total exams for FY 80-82 464 ND 4 I											
Department of the Treasury:															
Secret Service	NA			46.8	46.9	4.5	6.8	3 year average							
Bureau of Alcohol, Tobacco and Firearms	3:49	3:47	3:37	51.1	37.5	7.4	4.0	40.2	47.2	5.5	7.1	28.5	57.4	6.3	7.8
U.S. Postal Service	1.86 hours average			11	83	4	1.0	10	83	4	2	17.3	73.1	6.8	1.2

NOTES: D = Deceptive
ND = Nondeceptive
I = Inconclusive
NO = No Opinion
NA = Not available

Table B-4.—Long-term Confirmation of Exam Results (percent confirmed)

	Fiscal year 1980		Fiscal year 1981		Fiscal year 1982 (incomplete)	
	D	ND	D	ND	D	ND
Department of Defense:						
Army Criminal Investigation Command	75.4	20.9	68.8	20.9	72.8	49.3
	D confirmation via confession, court conviction, legal determination ND confirmation via legal determination or location of other suspect					
Army Intelligence and Security Command	70		82		83	
	Confirmation primarily via examinee admission					
Navy	42		45		46	
Air Force	50	10	3 year average			
	Confirmation by other evidence					
Marines			24		25	
National Security Agency	Data not available					
Department of Justice:						
Federal Bureau of Investigation	1,966 of 3,527 deceptives confirmed by confession					
Drug Enforcement Administration	65% of deceptives confess during post-test interrogation 85%of ND confirmed by subsequent investigations					
Department of the Treasury:						
Secret Service	Over 90% of opinions are confirmed 70% of D confirmed by admissions or confessions					
Bureau of Alcohol, Tobacco and Firearms	Data not available					
U.S. Postal Service	43% of D confessed		36% of D confessed		39% of D confessed	

D = Deceptive
ND = Nondeceptive

vestigation Command, Army Intelligence Command, and Secret Service. Confirmation of deceptive exam results is primarily by examinee admissions or confessions. Confirmation of nondeceptive exam results is generally more difficult, with nondeceptive confirmation rates of less than 50 percent indicated by all agencies reporting except DEA.

Use of Polygraph Examination Results

In general, with the exception of NSA, polygraph test results are used as an investigatory tool in specific criminal, counterintelligence, intelligence, or personnel security cases. Polygraph examinations are voluntary in the sense that agencies in general are proscribed from forcing individuals to take an examination, or from penalizing or taking adverse action against individuals who refuse to take an examination. However, at NSA, where a polygraph examination is part of the preemployment security screening process for all job applicants, refusal to take polygraph examination may result in failure to be accepted for employment. Also, the FBI noted that in cases where an FBI employee is asked to take a polygraph examination but refuses, the refusal may lead to an adverse inference being drawn.

Overall, agencies were not able to provide specific information on how the results of polygraph exams were actually used, since the agency office conducting the examination is usually different from the office conducting the investigation and taking action. Statistics on use of examination results apparently are not maintained, at least not on a centralized basis. Also, the results of a polygraph examination are usually only one of several sources of information relevant to a specific investigation. In fact, agency regulations generally require that polygraph results "be used selectively as an investigative aid" and not "to the exclusion of other evidence or knowledge obtained during the course of a complete investigation" (FBI regulation 13-22.2(2), 1981).

Federal Agency Polygraph Research

Based on information provided by Federal agencies, the major past, present, and future Federal polygraph research is summarized in table B-5. Research on the polygraph instrument itself includes a 1966-67 calibration study (U.S. Army), a 1966-67 technical evaluation study (Navy under contract to National Bureau of Standards), 1969-70 and 1975-77 cardioactivity monitor studies (Air Force), a current cardioactivity monitor study (FBI), and the current 10-year instrumentation research sponsored by the Army Criminal Investigation Command and Army Security and Intelligence Command and intended to develop a new polygraph instrument utilizing state-of-the-art technology. Research on polygraph validity and reliability, broadly defined, includes a 1962 validity study (Air Force), a 1965-67 reliability study (Army Criminal Investigation), 1979-81 counterintelligence screening test validity study (Army Intelligence), and the planned 1984-85 validity and reliability study cosponsored by the FBI and Secret Service. Also, in 1976-78, the Department of Justice sponsored validity and reliability studies by university researchers David Raskin and David Lykken. Finally, both Army Intelligence and NSA are planning research on polygraph countermeasures.

Table B-5.—Selected Federal Agency Polygraph Research

	Past polygraph research		Present/future research	
Department of Defense:				
Army Criminal Investigation Command	1965-67	Validation study of polygraph examiner judgments (reliability study known as "Bersh" study	1981-90	Instrument research and development project (to develop a state-of-the art polygraph instrument)
	1966-67	Calibration study of polygraph instrument		
	1973	Comparison of voice analysis and polygraph (U.S. Army Land Warfare Laboratory)		
Army Intelligence Command	1979-81	Validation and reliability study of counterintelligence screening test	1981-90	Instrument research and development project
Navy	1966-67	Technical evaluation study of polygraph instrument (under contract to National Bureau of Standards)	None	Planning research on polygraph countermeasure and anticountermeasures
Air Force	1962	Polygraph validity study	None	
	1965	Analysis of polygraphic data		
	1969-70	Development and validation studies of cardioactivity monitor		
	1975-78	Reliability and validity studies of cardioactivity monitor		
Marines		None	None	
National Security Agency	1983	Review of scientific literature on polygraph validity, reliability and utility	1983-84	$65,000 pilot study of effect of drugs/hypnosis/nonverbal techniques on polygraph validity
Department of Justice:				
Federal Bureau of Investigation		None	1984-85	Polygraph validity and reliability research (criminal investigatory context)
			1984	Instrumentation research (on monitoring blood pressure)
Law Enforcement Assistance Administration	1976-78	Raskin and Lykken studies of polygraph validity and reliability		
Department of the Treasury:				
Secret Service		Participated in Raskin study		Cooperation with planned FBI study on polygraph validity and reliability
Bureau of Alcohol, Tobacco and Firearms		None		None
U.S. Postal Service		None		None
Department of State		None		None
Office of Personnel Management		None		None
Central Intelligence Agency		Classified research		Data not available

Appendix C

Coding Form

DRAFT CODING FORM 7/18/83

Coder__________

AUTHOR____________________________________
YEAR_____

STUDY ID____________________
OUTCOME NO._________________
TOTAL NUMBER OF OUTCOMES IN THIS ANALYSIS____________

TYPSTUD1, Type of study: analog or field	(1) analog, (2) field
TYPESTU2, Type of study:	(1) detection, (2) blind evaluation of charts, (3) judgment of accuracy based on other criteria, (4) "utility study," (5) judgment of accuracy based in pg and other criteria

SUBJECTS

NSUBJS, Number of subjects or cases	_____
TYPSUBJS, Type of subj pop	(1) college students, (2) general pop, (3) non-crim. military personnel, (4) non-military criminals or suspects, (5) military criminals or suspects, (6) police informants, (7) prison inmates, (8) police applicants, (9) private employment applicants, (10) gov't employees or applicants, (11) victims, (12) witnesses
*CASESRC, Source of cases for judgment	(1) polygraph school files, (2) police files, (3) military files
PCTMALE,	_____
PURPOSE	(1) pre-employment, (2) crim investigation

Polygraph Coding Form
Page 2

POLYGRAPH CHARACTERISTICS

BASERATE, Base rate of guilt	_____
GROUNDME, Method of establishing ground truth	(1) Majority judgment, (2) Unanimous judgment, (3) Confession, (4) Court decision, (5) Mock crime or contrived story, (6) real crime "set up" by experimenter, (7) not verified, (8) not specified
ACCUR, Experimenter's judgment of accuracy of basis for ground truth (see Barland, 1982)	(1) low, (2) high
QUESDES, Method for designing control questions or pretest interview	(1) Standard for all Ss; (2) Customized
TECHNIQU, Type of question technique	(1)ZOC, (2) MGQT, (3) POT, (4) ZOC & MGQT, (5) ZOC & POT, (6) MGQT & POT, (7) GQT, (8) ZOC & GQT, (9) MGQT & GQT, (10) POT & GQT, (11) GK, (12) RI, (13) RCQT
STIM, Stim test included?	(1) Yes, (2) No
MACHINE, Machine type	(1) Lafayette 4 channel Model 76058, (2) Narco Bio-system polygraph, (3) 3 channel Stoelting, (4) 4 channel Stoelting, (5) Sanborn 150 Recorder, (6) Keeler polygraph, (7) Stoelting with CAM, (8) Grass Model 7, (9) physiograph, (10) 5 Channel Reid, (11) varied ___________, (12)__________________, (13)____________________
PASTE, Type of contact paste	(1) Sanborn, (2) Beckman, (3) NaCl w/cornstarch

Polygraph Coding Form
Page 3

PHYSMEAS, Phys. measure used for results	(1) SCR/GSR, (2) Respiration, (3) Blood pressure, (4) Heart rate, (5) Cardiovascular unpsecified, (6) finger pulse volume, (7) some combination
PHYSME2, Were other phys. measures taken and not used in analysis?	(1) Yes, (2) No
PHYSME3, If ans. to PHYSME2 is yes, why?	(1) results inconclusive, (2) not given
CHARTS, Number of charts on which examiners' judgment based	_____
PROCED, Did procedure differ from standard in any way (e.g., Podlesny & Raskin did not review control questions with Ss)	(1) Yes, (2) No
PROCVARY, Way procedure varied	______ (Use variable list code#)
EXAMEQ, Did examiners do own init. ratings (i.e., chart interpretations)	(1) Yes, (2) No
If answer to EXAMEQ is "No," answer following with respect to those who did do init. ratings (Note these are not ultimate judges in field studies)	
**PGBLIND, Were raters blind to subj condition?	(1) Yes, (2) No
**KNOWRATR, Did raters know rate of guilt?	(1) Yes, (2) No
**RATEXPF, Raters exp. ranged from (in yrs.)	_____
**RATEXPT, Raters exp. ranged to (in yrs.)	_____
OBJRAT, Was orig. rating objective?	(1) high (specific measurement of phys. variables), (2) medium (score assigned to subjective assessment, (3) low (rating of guilt or innocence based on visual assessment, (4) very low (rating of guilt or innocence based on case files, clinical assessment etc.)
NEXAM, Number if initial examiners	__________

Polygraph Coding Form
Page 4

INCZONE, Inconclusive zone (+ or - x)	_____
PGPCTMAL, % of Polygraphers Male	_____
PGEXP, Avg. yrs Poly training and experience	_____
EXAMEXPF, Examiners' exp. range from (in yrs.)	_____
EXAMEXPT, Examiners' exp. ranged to	_____
PGTYPE, Type of initial examiner	(1) private, (2) police, (3) military, (4) other govt, (5) trainees, (6) not a prof. examiner
PGTRN, Place polygraph examiner trained	(1) Reid, (2) Army, (3)
*JUDGES, Judge characteristics	(1) Polygraphers trained at same school, (2) Polygraphers trained at different school, (3) law enforcement agents, (4) legal professionals (lawyers, judges), (5) Same as initial examiners ("utility" studies), (6) Statistical analysis, (7) Other methods of ideniification (fingerprints, handwriting, eyewitness), (8) Other, (a) Polygraphers [other than (1) & (2)]
NJUDGES, Number of judges or evaluators (not initial examiners)	_____
KNOWRATJ, Did judges know base rate of guilt?	(1) Yes, (2) No
JUDGAGRE, Method of judge agreement (if panel)	(1) Unanimous, (2) Majority
*JUDGEXPF, Judges exp. ranged from (in yrs.)	_____
*JUDGEXPT, Judges exp. ranged to (in yrs.)	_____
*JUDGEXP2, Judges exp. ranged to	(1) less than 1 yr., (2) greater than 1 yr.
*JUDGEXP3, Judges exp. ranged from	(1) less than 1 yr., (2) greater than 1 yr.
*AVJUDEXP, Av. judge exp (yrs.)	_____

Polygraph Coding Form
Page 5

DESIGN

SAMPLING, Random selection of Ss or cases?	(1) Yes, (2) No
EXCLU, If not randomly selected, % of population not included in sample	_____
BASISSEC, Basis of selection (use variable code listing)	_____
ATTRIT, % attrition from sample	_____
BASISATT, Basis of attrition (use variable code listing	_____
KNOWRATE, Did init. examiners know rate of guilt?	(1) Yes, (2) No
MOTIV, Were subjects offered inducement to beat machine? (analogue only)	(1) Yes, (2) No
PGBLIND2, Did examiners know Ss were in an exp?	(1) Yes, (2) No
INDEPEND, Was initial polygraph rating blind (independent of examination?)	(1) Yes, (2) No
*OBJRAT2, Were "judges" ratings objective?	(1) high (specific measurement of phys. variables), (2) medium (score assigned to subjective assessment, (3) low (rating of guilt or innocence based on visual assessment, (4) very low
FACTOR1A, Factorial effect tested (use variable code listing)	_____
FACTOR1B, Was factorial effect 1A significant?	(1) Yes, (2) No
FACTOR2A, Second factorial effect tested?	_____
FACTOR2B, Was factorial effect 2A significant	(1) Yes, (2) No
FACTOR3A, Third factorial effect tested?	_____
FACTOR3B. Was factorial effect 3B signifiicant?	(1) Yes, (2) NO
FACTOR4A, Fourth factorial effect tested?	_____
FACTOR4B, Was factorial effect 4A significant?	(1) Yes, (2) No

OUTCOME

DETECTION STUDIES

GC ______
GNC ______
GIN ______
IC ______
INC ______
IIN ______

UNIT, Unit of analysis for outcome 1) Persons
(2) Questions,

JUDGMENT STUDIES

JGC ______
JGNC ______
JGIN ______
JIC ______
JINC ______
JIIN ______

OTHER CROSS-VALIDATION STUDIES

GC2 ______
GNC2 ______
GIN2 ______
IC2 ______
INC2 ______
IIN2 ______

CONTINUOUS SCORES (Means and signif. tests)

GUILTY, Mean for guilty (deceptive) subjects ______
INNO, Mean for innocent (truthful) subjects ______
SIGTEST, Significance test used (1) F, (2) t
(3) , (4)
SIGDIFF, Was difference significant? (1) Yes, (2) No

Appendix D

Acronyms and Glossary

Acronyms

ANS —autonomic nervous system
CIA —Central Intelligence Agency
CQT —control question technique
DCI —Director of Central Intelligence
DLCQ —directed lie control question [technique]
DOD —Department of Defense
DOJ —Department of Justice
EDR —electrodermal response
GKT —guilty knowledge test
GSR —galvanic skin response
GQT —general question test
JAG —Judge Advocate General
LEAA —Law Enforcement Assistance Administration
MGQT —modified general question test
MMPI —Minnesota Multiphasic Personality Inventory
NSA —National Security Agency
NSDD —National Security Decision Directive
OPM —Office of Personnel Management
Pd —psychopathic deviate [scale]
POT —peak of tension [test]
RCQT —Reid control question technique
R/I —relevant/irrelevant [technique]
SCI —Sensitive Compartmented Information
SCR —skin conductance response
USAMPS—U.S. Army Military Police School
USPS —U.S. Postal Service
ZOC —zone of comparison [technique]

Glossary

analog studies: Analog studies are laboratory studies of polygraph testing that simulate actual field examinations. Typical components of field examinations are replicated. The goal of such studies is to test the validity of various polygraph techniques under controlled conditions.

aperiodic checking: Polygraph tests conducted at irregular times with randomly or otherwise selected personnel to ask questions for internal security purposes.

autonomic lability: Term used to describe individual differences in autonomic arousal.

base rate: The number of guilty (or innocent) subjects as a percentage of the total.

baseline: The readings on a polygraph chart that form a point of comparison for the physiological responses to the polygraph questions.

classified information: Information that pertains to national security and by definition, cannot be disclosed to others without clearance.

clinical components: Components of a polygraph test procedure, including "proper" examiner attitude and relationship with subjects, that attempt to ensure accuracy.

construct validity: The extent to which a test or procedure measures what it is designed to measure.

control question technique: A polygraph question technique that incorporates control questions which are designed to be arousing for nondeceptive subjects and less arousing for deceptive subjects than the relevant questions.

counterintelligence: Efforts of an organization to stop outside groups from gaining information about itself.

counterintelligence screening examinations: Examinations given to personnel who already have access to classified information.

electrodermal response: A physiological measure that has been shown to be related to physiological arousal. It is measured as the electrical resistance of the skin through the use of electrodes attached to the fingertips.

external validity: The established generalizability of a study to particular subject populations and settings.

false negative: An erroneous decision that an individual is not deceptive when she or he is actually deceptive.

false positive: An erroneous decision that a person is being deceptive when he or she is actually being truthful.

field testing: Actual techniques used by polygraph examiners.

generalizability: The extent to which results of previous investigations can be used in evaluation of present investigations.

ground truth: The establishment of actual guilt or innocence. In a field study it is based on a criterion independent of the polygraph test (e.g., confession, judicial outcome, panel decision).

inconclusives: Outcome of an examination in which it cannot be determined from the subject's responses whether he or she is deceptive or nondeceptive.

interaction: An occurrence which affects validity of polygraph testing because individual character traits or situational factors might result in unexpected physiological responses.

internal validity: The degree to which a study has controlled for extraneous variables which may be related to the study outcome.

irrelevant questions: Neutral questions designed to assess the subject's baseline physiological response to questioning and to provide a rest between relevant questions.

numerical scoring: The assignment of numbers to polygraph chart responses.

physiological arousal: Responses related to increases in anxiety. Those measured in polygraph examinations include electrodermal response, blood pressure, and respiration rate.

polygraph chart: A continuous graph on which a subject's physiological responses are registered.

predictive association: An index which measures the proportional reduction in the probability of error in predicting one category (in this case, deception) when the second category (in this case, polygraph examination results) is known.

predictive validity: The accuracy with which criterion scores obtained in the future can be estimated from test data obtained in the present.

preemployment screening: The use of polygraph testing to question employee applicants.

pretest interview: The first portion of the polygraph testing procedure in which subjects are informed about the examination and their rights. In some pretest interviews, examiners also make observations about subjects' behavior to assist in determinations of deceptiveness or nondeceptiveness.

psychopathy: A psychiatric diagnostic category signifying a character style prone to criminal activity and amoral, manipulative behavior.

random sampling: A procedure used to obtain representative samples from a population. In complete random sampling, each subject in the population must have an equal chance of being selected and the selection or nonselection of one subject cannot influence the selection or nonselelction of another.

relevant/irrelevant technique: An examination technique that utilizes two types of questions: relevant questions and neutral questions intended to assess the subject's baseline response.

relevant questions: Polygraph test questions about the topic or topics under investigation.

reliability: The degree to which a test yields repeatable results. Reliability also refers to consistency across examiners/scorers.

Sensitive Compartmented Information: Classified information above the top secret level.

socialization: The process in and by which individuals learn the ways, ideas, beliefs, values, patterns, and norms of a particular culture and adapts them as a part of their own personalities.

validity: A measure of the extent to which an observed situation reflects the "true" situation.

References

References

1. Abrams, S., "Polygraph Validity and Reliability: A Review," *Journal of Forensic Sciences* 17:313-327, 1973.
2. *Addington* v. *Texas*, 441 US 418, 1979.
3. American Polygraph Association, *Manual for Polygraph School Inspections* (Lithicum Heights, Md.: American Polygraph Association, 1983).
4. American Polygraph Association, "Polygraph School Accreditation Standards," 1983.
5. American Psychological Association, *Standards for Educational and Psychological Tests* (Washington, D.C.: American Psychological Association, 1974).
6. American Psychological Association, *Joint Technical Standards for Educational and Psychological Testing*, unpublished draft, Washington, D.C., 1983.
7. Ansley, N., "A Review of the Scientific Literature on the Validity, Reliability and Utility of Polygraph Techniques" (Ft. Meade, Md.: National Security Agency, 1983).
8. Ansley, N., *Quick Reference Guide to Polygraph Admissibility, Licensing Laws, and Limiting Laws* (Severna Park, Md.: American Polygraph Association, 1983).
9. Ansley, N., Horvath, F., and Barland, G. H., *Truth and Science: A Bibliography*, 2d ed. (Lithicum Heights, Md.: American Polygraph Association, 1983). A comprehensive index to international literature on the detection of deception and the polygraph (lie detector) technique.
10. Backster, C., "Methods of Strengthening Our Polygraph Technique," *Police* 6:61-68, 1962.
11. Backster, C., *Standardized Polygraph Notepack and Technique Guide* (San Diego, Calif.: C. Backster, 1979).
12. Balloun, K. D., and Holmes, D. S., "Effects of Repeated Examinations on the Ability to Detect Guilt With a Polygraphic Examination: A Laboratory Experiment With a Real Crime," *Journal of Applied Psychology* 64:316-322, 1979.
13. *Barefoot* v. *Estelle*, SILW 5190, June 28, 1983.
14. Barland, G. H., "A Survey of the Effect of the Polygraph in Screening Utah Job Applicants: Preliminary Results," in: U.S. Congress, Senate Judiciary Committee, hearings, 1978.
15. Barland, G. H., "An Introduction to the Number Test," *Polygraph* 7:203-208, 1978.
16. Barland, G. H., "A Validation and Reliability Study of the Counterintelligence Screening Test," unpublished manuscript, Security Support Battalion, Military Intelligence Group, U.S. Army, Fort George Meade, Md., 1981.
17. Barland, G. H., "On the Accuracy of the Polygraph: An Evaluative Review of Lykken's Tremor in the Blood," *Polygraph* 11:258-272, 1982.
19. Barland G. H., "Detection of Deception Theory," unpublished, November 1982.
20. Barland, G. H., and Raskin, D. C., "Detection of Deception," *Electrodermal Activity in Psychological Research*, W. F. Prokasy and D. C. Raskin (eds.) (New York: Academic Press, 1973), pp. 418-471.
21. Barland, G. H., and Raskin, D. C., "An Evaluation of Field Techniques in Detection of Deception," *Psychophysiology* 12:321-330, 1975.
22. Barland, G. H., and Raskin, D. C., "Validity and Reliability of Polygraph Examinations of Criminal Suspects," report No. 76-1, contract No. 75-N1-99-0001 (Washington, D.C.: National Institute of Justice, Department of Justice, 1976).
24. Barton, M., "A Study of the Admissions Made During Pre-Employment Polygraph Examinations (1964-1975) and Their Significance to the Business Community," unpublished masters thesis, Sam Houston State University, 1976.
25. Belt, J. A., and Holden, P. B., "Polygraph Usage Among Major U.S. Corporations," *Personnel Journal* 80-86, 1978.
26. Ben-Ishai, A., "Some Remarks on Polygraph Research," paper presented at the Ninth Annual Meeting of the American Academy of Polygraph Examiners, Chicago, Ill., August 1962.
27. Ben-Shakhar, G., Bar-Hillel, M., and Lieblich, I., *Trial by Polygraph: Scientific and Juridical Issues in Lie Detection*, Hebrew University, 1983.
28. Ben-Shakhar, G., Lieblich, I., and Bar-Hillel, M., "An Evaluation of Polygraphers' Judgements: A Review From a Decision Theoretic Perspective," *Journal of Applied Psychology* 67:701-713, 1982.
29. Bersh, P. J., "A Validation Study of Polygraph Examiner Judgements," *Journal of Applied Psychology* 53:399-403, 1969.
30. Bitterman, M. E., and Marcuse, F. L., "Cardiovascular Responses of Innocent Persons to Criminal Interrogation," *American Journal of Psychology* 60:407-412, 1947.
32. Blaylock, K. D., "Use of Polygraphs and Prepublication Review," testimony before joint hearing of House Post Office and Civil Service

Subcommittee on Civil Service and House Judiciary Subcommittee on Civil and Constitutional Rights, Apr. 21, 1983.

33. Blum, R. H., and Osterloh, W., "The Polygraph Examination as a Means for Detecting Truth and Falsehood in Stories Presented by Police Informants," *The Journal of Criminal Law, Criminology and Police Science* 59:133-137, 1968.
34. Borchard, E., *Convicting the Innocent* (Garden City, N.Y.: Garden City Publishing Co., 1932).
35. Bradley, M. T., and Janisse, M. P., "Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures," *Psychophysiology* 18: 307-314, 1981.
36. Brown, C. C., *Methods in Psychophysiology* (Baltimore: The Williams & Wilkins Co., 1967).
37. Buckley, J., letter to Fred B. Wood, Office of Technology Assessment, 1983.
39. Carlucci, F. C., "Degradation of Operational Readiness/Mission Accomplishment Due to Personnel Security Investigation Shortfalls," memorandum, U.S. Department of Defense, Aug. 6, 1982.
40. Cimmerman, A., "They'll Let Me Go Tomorrow," *Criminal Defense* 8, 7-10, 1981.
41. Cook, T. D., and Campbell, D., *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Chicago: Rand McNally, 1979).
42. Corcoran, J. F. T., Lewis, M. D., and Garver, R. B., "Biofeedback Conditioned Galvanic Skin Response and Hypnotic Suppression of Arousal: A Pilot Study of Their Relation to Deception," *Journal of Forensic Sciences* 23:155-162, 1978.
43. Correa, E. I., and Adams, H. E., "The Validity of the Pre-employment Polygraph Examination and the Effects of Motivation," *Polygraph* 10:143-156, 1981.
44. Cureton, R., "A Consensus as to the Validity of Polygraph Procedures," *Tennessee Law Review* 22:18-32, 1953.
45. Cutrow, R. J., Parks, A., Lucas, N., and Thomas, K., "The Objective Use of Multiple Physiological Indices in the Detection of Deception," *Psychophysiology* 9:578-588, 1972.
46. Davidson, P. O., "Validity of the Guilty-Knowledge Technique: The Effects of Motivation," *Journal of Applied Psychology* 52:62-65, 1968.
47. Davidson, W. A., "Validity and Reliability of the Cardio Activity Monitor," *Polygraph* 8: 104-111, 1979.
48. Davis, R. C., "Physiological Responses as a Means of Evaluating Information," *The Manipulation of Human Behavior,* A. D. Biderman and H. Zimmer (eds.) (New York: Wiley, 1961), pp. 142-169.
49. Dawson, M. E., "Physiological Detection of Deception: Measurement of Responses to Questions and Answers During Countermeasure Maneuvers," *Psychophysiology* 17:8-17, 1980.
50. Decker, R. E., "The Army Stimulation Test: A Control Procedure," *Polygraph* 7:176-178, 1978.
51. Edel, E. C., and Jacoby, J., "Examiner Reliability in Polygraph Chart Analysis: Identification of Physiological Responses," *Journal of Applied Psychology* 60:632-634, 1975.
52. Edwards, R. H., *A Survey: Reliability of Polygraph Examination Conducted by Virginia Polygraph Examiners,* report for the Commonwealth of Virginia (Richmond, Va.: Bureau of Forensic Science, Department of General Services, 1981).
53. Ekman, P., Levenson, R. W., and Friesen, W. V., "Autonomic Nervous System Activity," *Science* 221:1208-1210, 1983.
54. Elaad, E., and Schahar, E., "Polygraph Field Validity," *Scientific Interrogation in Criminal Investigation,* I. Nachson (ed.), selected papers presented at the First National Conference on Scientific Interrogation in Criminal Investigation at Bar-Ilan University and Ramat-Gan, Israel, November 1976.
55. Ennis, B. J., and Litwack, T. R., "Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom," *California Law Review* 62:693-752, 1974.
56. Ferguson, R. J., *The Polygraph in Private Industry* (Springfield, Ill.: Charles C. Thomas, 1966).
57. Fingerhut, K. R., "Use of the Stimulation Test in Preemployment Testing," *Polygraph* 7:185-188, 1978.
58. *Frye* v. *United States,* 293 F. 1013 (DC Cir. 1923).
59. Gatchel, R. J., Smith, J. E., Kaplan, N. M., et al., "The Effect of Propanolol on Polygraphic Detection of Deception," unpublished manuscript, 1983.
60. Giannelli, P. C., "The Admissibility of Novel Scientific Evidence: *Frye* v. United States, a Half Century Later," *Columbia Law Review* 80: 1198-1250, 1980.
61. Giesen, M., and Rollison, M. A., "Guilty Knowledge Versus Innocent Associations: Ef-

fects of Trait Anxiety and Stimulus Context on Skin Conductance," *Journal of Research in Personality* 14:1-11, 1980.

62. Ginton, A., Netzer, D., and Elaand, E., "A Method for Evaluating the Use of the Polygraph in a Real-Life Situation," *Journal of Applied Psychology* 67:131-137, 1982.
63. Glass, G. V., McGaw, B., and Smith, M. L., *Meta-Analysis in Social Research* (Beverly Hills, Calif.: Sage Publications, 1982).
64. Goodman, L. A., and Kruskal, W. H., "Measures of Association for Cross-Classification," *Journal of the American Statistical Association* 49:732-764, 1954.
65. Gustafson, L. A., and Orne, M. T., "Effects of Heightened Motivation on the Detection of Deception," *Journal of Applied Psychology* 43: 408-411, 1963.
66. Gustafson, L. A., and Orne, M. T., "The Effects of Task and Method of Stimulus Presentation on the Detection of Deception," *Journal of Applied Psychology* 49:412-417, 1964.
67. Gustafson, L. A., and Orne, M. T., "Effects of Perceived Role and Role Success on the Detection of Deception," *Journal of Applied Psychology* 49:412-417, 1965.

67a. Hammond, D. L., "The Responding of Normals, Alcoholics and Psychopaths in a Laboratory Lie-Detection Experiment," unpublished doctoral dissertation, California School of Professional Psychology, 1980.

68. Harnon, E., "Evidence Obtained by Polygraph: An Israeli Perspective," *The Criminal Law Review* 340-349, 1982.
69. Harrelson, L. H., *The Keeler Technique,* 2d ed. (Chicago: Keeler Polygraph Institute, 1964).
71. Hayden, T., "ACLU Project Privacy," testimony of Trudy Hayden, in U.S. Congress, Senate Judiciary Committee, hearings, 1978.
72. Hayden, T., "Employers Who Use Lie Detector Tests," *Business & Society Review* 41:16-21, 1982.
73. Hays, W. L., *Statistics,* 3d ed. (New York: Holt, Rinehart, & Winston, 1981).
74. Heckel, R. V., Brokaw, J. R., Salzberg, H. C., and Wiggins, S. L., "Polygraphic Variations in Reactivity Between Delusional, Nondelusional, and Control Groups in a Crime Situation," *Journal of Criminal Law, Criminology and Police Science* 53:380-383, 1962.
75. Honts, C. R., and Hodes, R. L., "The Effect of Simple Physical Countermeasures on the Detection of Deception," paper presented at meetings of the Society for Psychophysiological Research, Minneapolis, Minn., 1982.
76. Honts, C. R., and Hodes, R. L., "The Effects of Multiple Physical Countermeasures on the Detection of Deception," *Psychophysiology* 19:564-565 (abstract), 1982.
77. Honts, C. R., and Hodes, R. L., "The Effect of Simple Physical Countermeasures on the Detection of Deception," *Psychophysiology* 19:564 (abstract), 1982.
78. Honts, C. R., Raskin, D. C., and Kircher, J. C., "Detection of Deception: Effectiveness of Physical Countermeasures Under High Motivation Conditions," paper presented at meetings of the Society for Psychophysiological Research, September 1983.
79. Horvath, F. S., "The Police Candidate Polygraph Examination: Considerations for the Police Administrator," *Police* 33-39, 1972.
80. Horvath, F. S., "Verbal and Nonverbal Clues to Truth and Deception During Polygraph Examinations," *Journal of Police Science and Administration* 1:138-152, 1973.
81. Horvath, F. S., "Detection of Deception: A Review of Field and Laboratory Research," *Polygraph* 5:107-145, 1976.
82. Horvath, F. S., "The Effect of Selected Variables on Interpretation of Polygraph Records," *Journal of Applied Psychology* 62:127-136, 1977.
83. Horvath, F. S., "Detecting Deception in Eyewitness Cases: Problems and Prospects in the Use of the Polygraph," *Advances in the Psychology of Eyewitness Testimony,* J. Wells and B. Loftus (eds.) (New York: Cambridge University Press), in press.
84. Horvath, F. S., and Reid, J. E., "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception," *The Journal of Criminal Law, Criminology and Police Science* 62:276-281, 1971.
85. Hunter, F. L., and Ash, P., "The Accuracy and Consistency of Polygraph Examiners' Diagnoses," *Journal of Police Science and Administration* 1:370-375, 1973.
86. Iacono, W. G., Boisvenu, G. A., and Fleming, J. A., "The Effects of Diazepam and Methylphenidate on the Electrodermal Detection of Guilty Knowledge," unpublished manuscript, University of British Columbia, Vancouver, Canada.
87. Jones, E. E., and Sigall, H., "The Bogus Pipeline: A New Paradigm for Measuring Affect

and Attitude," *Psychological Bulletin* 76: 349-364, 1971.

89. Keeler, L., "Debunking the 'Lie Detector,'" *The American Journal of Police Science* XXV:153-159, 1934.

89a. Kircher, J. C., "Computerized Decision-Making and Patterns of Activation in the Detection of Deception," unpublished doctoral dissertation, University of Utah, 1983.

90. Kircher, J. C., and Raskin, D. C., "Computerized Decisionmaking in the Detection of Deception," *Psychophysiology* 18:204-205, 1981.

91. Kircher, J. C., and Raskin, D. C., "Clinical Versus Statistical Lie Detection Revisited: Through a Lens Sharply," unpublished, Department of Psychology, University of Utah, Salt Lake City, Utah, 1983.

92. Kleinmuntz, B., and Szucko, B., "On the Fallibility of Lie Detection," *Law and Society Review* 17:84-104, 1982.

93. Kubis, J. F., *Studies in Lie Detection: Computer Feasibility Considerations* contract No. AF-30-(602)-2270, project No. 5534 submitted to U.S. Air Force, available from the Armed Services Technical Information Service (AD-284902), Arlington, Va. 1962.

94. Kubis, J. F., *Analysis of Polygraphic Data*, Technical Report No. RADC-TDR-64-101, January 1965.

95. Kugelmass, S., "Reactions to Stress," contract No. AF-61-(052)839, U.S. Air Force Office of Scientific Research, 1967.

96. Kuglemass, S., Lieblich, I., and Bergman, Z. "The Role of 'Lying' in Psychophysiological Detection," *Psychophysiology* 3:312-315, 1967.

97. Kugelmass, S., and Lieblich, I., "The Relation Between Ethnic Origin and GSR Reactivity in Psychophysiological Detection," *Journal of Applied Psychology* 52:158-162, 1968.

98. Kugelmass, S., Lieblich, I., Ben-Ishai, A., Opatowski, A., and Kaplan M., "Experimental Evaluation of Galvanic Skin Response and Blood Pressure Change Indices During Criminal Interrogation," *The Journal of Criminal Law, Criminology and Police Science* 59:632-635, 1968.

100. Lahri, S. K., and Ganguly, A. K., "Experimental Study of the Accuracy of Polygraph Technique in Diagnosis of Deception With Volunteer and Criminal Subjects," *Polygraph* 7:89-94, 1978.

101. Larson, J. A., *Lying and Its Detection* (Chicago: University of Chicago Press, 1932).

103. Larson, J. A., "Deception Tests and Lie Detectors," *Journal of Criminal Law and Criminology* 12:390-399, November 1921.

104. Lazarus, R. S., *Psychological Stress and the Coping Process* (New York: McGraw Hill, 1966).

105. Lykken, D. T., "The GSR in the Detection of Guilt," *Journal of Applied Psychology* 43:385-388, 1959.

106. Lykken, D. T., "The Validity of the Guilty Knowledge Technique: The Effects of Faking," *Journal of Applied Psychology* 44:258-262, 1960.

107. Lykken, D. T., "Psychology and the Lie Detector Industry," *American Psychologist* 29:725-739, 1974.

108. Lykken, D. T., *A Tremor in the Blood: Uses and Abuses of the Lie Detector* (New York: McGraw Hill, 1981).

109. Lykken, D. T., "The Psychopath and the Lie Detector," *Psychophysiology* 15:137-142, 1978.

110. Lykken, D. T., "The Detection of Deception," *Psychological Bulletin* 86:47-53, 1979.

111. Lyon, V. W., "Deception Tests With Juvenile Delinquents," *Journal of Genetic Psychology* 48:494-497, 1936.

112. Macy J. W., Jr., Chairman, U.S. Civil Service Commission, letter to Hon. John E. Moss, Chairman, Foreign Operations and Government Information Subcommittee, House Committee on Government Operations, Nov. 22, 1965.

113. MacNitt, R. D., "In Defense of the Electrodermal Response and Cardiac Amplitude as Measures of Deception," *Journal of Law and Criminology* 33:266-275, 1942.

114. Marston, W. M., "Systolic Blood Pressure Changes in Deception," *Journal of Experimental Psychology* 2:143-163, 1917.

115. Matte, J. A., "The Art and Science of the Polygraph Technique" (Springfield, Ill.: Charles C. Thomas, 1980).

116. McEvoy, J. P., "The Lie Detector Goes Into Business," *The Reader's Digest*, January 1941, pp. 69-72.

118. Monahan J., *The Clinical Prediction of Violent Behavior* (Washington, D.C.: U.S. Government Printing Office, 1981

119. More, H. W., "Polygraph Research and the University," *Law and Order* 14:73-78, 1966.

120. Moroney, W. F., and Zenhausern, R. J., "Detection of Deception as a Function of Galvanic Skin Response Recording Methodology," *The Journal of Psychology* 80:255-261, 1972.

121. Mullenix, P. A., and Reid, J. A., "The Pretest

Interview and Its Role in the Detection of Deception," *Polygraph* 9:74-85, 1980.

122. Orne, M. T., "Implications of Laboratory Research for the Detection of Deception," *Legal Admissibility of the Polygraph*, N. Ansley (ed.) (Springfield, Ill.: Charles C. Thomas, 1975), pp. 94-119.
123. Orne, M. T., Thackray, R. I., and Paskewitz, D. A., "On the Detection of Deception: A Method for the Study of the Physiological Effects of Psychological Stimuli," *Handbook of Psychophysiology*, N. Greenfield and R. Sternbach (eds.) (New York: Holt, Rinehart, & Winston 1972), pp. 743-785.
124. Peters, R. B., "A Survey of Polygraph Evidence in Criminal Trials," *American Bar Association Journal* 68:162-165, 1982.
125. Pillemer, D. B., and Light, R. J., "Benefiting From Variation in Study Outcomes," *New Directions for Methodology of Social Behavior Science: Quantitative Assessment of Research Domains*, R. Rosenthal (ed.) (San Francisco: Jossey-Bass, 1980).
126. Podlesny, J. A., and Raskin, D. C., "Physiological Measures and the Detection of Deception," *Psychological Bulletin* 84.782-799, 1977.
127. Podlesny, J. A., and Raskin, D. C., "Effectiveness of Techniques and Physiological Measures in the Detection of Deception," *Psychophysiology* 15:344-358, 1978.
128. Privacy Protection Study Commission, *Personal Privacy in an Information Society* (catalog No. Y3.P93/5:1/9/77) (Washington, D.C.: U.S. Government Printing Office, July 1977).
130. Pyle, C. H., "The Defense Department's Polygraph Regulations of 1982," statement before the Subcommittee on Civil and Constitutional Rights, Committee on the Judiciary, U.S. House of Representatives, Jan. 25, 1982.
131. Quigley-Fernandez, B., and Tedeschi, J. T., "The Bogus Pipeline as Lie Dectector: Two Validity Studies," *Journal of Personality and Social Psychology* 36:247-256, 1978.
132. Raskin, D. C., personal communication, Aug. 23, 1983.
133. Raskin, D. C., *Reliability of Chart Interpretation and Sources of Errors in Polygraph Examinations* (report No. 76-3, contract No. 75-NI-99-0001, National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice, University of Utah, 1976.
134. Raskin, D. C. "Scientific Assessment of the Accuracy of Detection of Deception," *Psychophysiology* 15:143-147, 1978.
135. Raskin, D. C., "Science, Competence, and Polygraph Techniques," *Criminal Defense* 8:11-18, 1981.
136. Raskin, D. C., "The Scientific Basis of Polygraph Techniques and Their Uses in the Judicial Process," proceedings of the Stockholm Symposium on Witness Psychology, 1982, pp. 319-371.
137. Raskin, D. C., and Hare, R. D., "Psychopathy and Detection of Deception in a Prison Population," *Psychophysiology* 15:126-136, 1978.
138. Raskin, D. C., and Podlesney, J. A., "Truth and Deception: A Reply to Lykken," *Psychological Bulletin* 86:54-58, 1979.
139. Reid, J. E., and Inbau, F. E., *Truth and Deception—The Polygraph Technique*, 3d ed. (Baltimore: The Willliams & Wilkins Co., 1977).
140. Robbins, N. E., and Penley, W. J., "Review of Polygraph Charts on Nondeceptive Subjects," *Polygraph* 4:199-207, 1975.
141. Rosenthal, R., *Experimenter Effects in Behavioral Research* (New York: Irvington, 1976).
142. Rosenthal, R., and Rubin, D. B., "Comparing Effect Sizes of Independent Studies," *Psychological Bulletin* 2:500-504, 1982.
143. Rovner, L. I., Raskin, D. C., and Kircher, J. C., "Effects of Information and Practice on Detection of Deception," paper presented at the meetings of the Society for Psychophysiological Research, Wisconsin, 1979.
144. Sackett, P. R., and Decker, P. J., "Detection of Deception in the Employment Context: A Review and Critical Analysis," *Personnel Psychology* 32:487-507, 1979.
145. Saks, M. J., and Van Duizend, R., *The Use of Scientific and Technological Evidence in Litigation: Report and Indexed Bibliography* (Williamsburg, Va.: National Center for State Courts, 1983).
146. Sarbin, T. R., and Slagle, R. W., "Hypnosis and Psychophysiological Outcomes," *Hypnosis: Developments in Research and New Perspectives*, E. Fromm and R. E. Shor (eds.) (New York: Aldine, 1979).
147. Saxe, L., and Fine, M., *Social Experiments Methods for Design and Evaluation* (Beverly Hills, Calif.: Sage Publications, 1981).
148. Schroeder, J., "Zone Comparison Test Construction (ZCT) Summary Sheet," unpublished manuscript, 1983.
149. Senese, L., "Accuracy of the Polygraph Technique With and Without Card Test Stimulation," *Journal of Police Science and Administration* 4:274-276, 1976.
150. Sigall, H., and Page R., "Current Stereotypes:

A Little Fading, A Little Faking," *Journal of Personality and Social Psychology* 18:247-255, 1971.

151. Sigall, H., and Page, R., "Reducing Attenuation in the Expression of Interpersonal Affect via the Bogus Pipeline," *Sociometry* 35: 629-642, 1972.
152. Simon, M. A., "Shall We Ask the Lie Detector?" *Science, Technology, and Human Values* 8:3-13, 1983.
153. Skolnick, J. H., "Scientific Theory and Scientific Evidence: An Analysis of Lie Detection," *Yale Law Journal* 70:694-728, 1961.
154. Slowik, S. M., "Ethics in Pre-employment Polygraph Examinations," *Polygraph* 8:144-149, 1979.
155. Slowick, S. M., and Buckley, J. P., "Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure, and GSR Recordings," *Journal of Police Science and Administration* 3:305-309, 1975.
156. Smith, R., "Compilation of State and Federal Privacy Laws," *Privacy Journal* 1:79, 1981.
157. Smith, R., "Supplement to Compilation of State and Federal Privacy Laws 1982, 1983," *Privacy Journal.*
158. *Sternback, R. A., and Turskey, B., "Ethnic Differences Among Housewives on Psychophysical and Skin Potential Responses to Electric Shock," Psychophysiology* 1:241-246, 1965.
159. Szucko, J. J., personal communication, 1983.
160. Szucko, J. J., and Kleinmuntz, B., "Statistical Versus Clinical Lie Detection," *American Psychologist* 36:488-496, 1981.
161. Thackray, R. I., and Orne, M. T., "Effects of the Type of Stimulus Employed and the Level of Subject Awareness on the Detection of Deception," *Journal of Applied Psychology* 52:234-239, 1968.
162. Timm, H. W., "The Effect of Placebos and Feedback on the Detection of Deception," unpublished doctoral dissertation, College of Social Science, Michigan State University, 1979.
163. Timm, H. W., "Analyzing Deception From Respiration Patterns," *Journal of Political Science and Administration* 10:47-51, 1982.
164. Timm, H. W., "Effect of Altered Outcome Expectancies Stemming From Placebo and Feedback Treatment on the Validity of the Guilty Knowledge Technique," *Journal of Applied Psychology* 4:391-400, 1982.
165. U.S. Central Intelligence Agency, Director of Central Intelligence, "Investigative Scope and Adjudicative Procedures Among Intelligence Community Agencies: Personnel Security Survey," 1980.
166. U.S. Civil Service Commission, letter to the President, July 29, 1966.
167. U.S. Congress, House Committee on Government Operations, Subcommittee on Government Information and Foreign Operations, *Use of Polygraphs as "Lie Detectors" by the Federal Government,* 89th Cong., 1st sess., House Report No. 198 (Washington, D.C.: U.S. Government Printing Office, 1965).
168. U.S. Congress, House Committee on Government Operations, Subcommittee on Government Information and Foreign Operations, *Use of Polygraphs as Lie Detectors by the Federal Government (Part 2),* 89th Cong., 2d sess., House Report No. 280V (Washington, D.C.: U.S. Government Printing Office, 1966).
169. U.S. Congress, Senate Committee on the Judiciary, *Privacy, Polygraphs, and Employment,* a study prepared by the staff of the Subcommittee on Constitutional Rights (Washington, D.C.: U.S. Government Printing Office. 1974).
170. U.S. Congress, House Committee on Government Operations, *The Use of Polygraph and Similar Devices by Federal Agencies,* hearings before Subcommittee on Foreign Operations and Government Information, 93d Cong., 2d sess. (Washington, D.C.: U.S. Government Printing Office, 1974).
171. U.S. Congress, House Committee on Government Operations, *The Use of Polygraphs and Similar Devices by Federal Agencies,* Thirteenth Report by the Committee on Government Operations Together with Separate and Dissenting Views, 94th Cong., 2d sess. (Washington, D.C.: U.S. Government Printing Office, 1976).
172. U.S. Congress, Senate Committee on the Judiciary, *Polygraph Control and Civil Liberties Protection Act,* hearings before the Subcommittee on the Constitution (Washington, D.C.: U.S. Government Printing Office, 1978).
173. U.S. Congress, House Permanent Select Committee on Intelligence, Subcommittee on Oversight, *Security Clearance Procedures in the Intelligence Agencies* (Washington, D.C.: U.S. Government Printing Office, 1979).
174. U.S. Congress, House Committee on Education and Labor, *Pressures in Today's Workplace,* vol. III, hearings before the Subcommittee on Labor-Management Relations, Dec. 15, 1979 (Washington D.C.: U.S. Government Printing Office, 1980).

175. U.S. Congress, House Permanent Select Committee on Intelligence, *Preemployment Security Procedures of the Intelligence Agencies,* hearings before the Subcommittee on Oversight, May 16, 17, 24, June 21, 1979 (Washington, D.C.: U.S. Government Printing Office, 1980).

176. U.S. Department of Commerce, *The Cost of Crimes Against Business,* 1976.

177. U.S. Department of Defense, *Department of Defense Directive Number 5210.48, Conduct of Polygraph Examinations and the Selection, Training and Supervision of DOD Polygraph Examiners,* July 13, 1965.

178. U.S. Department of Defense, *Department of Defense Directive Number 5210.48, Polygraph Examinations and Examiners,* Oct. 6, 1975.

179. U.S. Department of Defense, Department of the Army, *Law Enforcement Investigations,* Army Field Manual 19-20 (Washington, D.C.: Department of the Army, 1977).

180. U.S. Department of Defense, *Select Panel Review of the Department of Defense Personnel Security Program,* Apr. 16, 1982.

181. U.S. Department of Defense, *Department of Defense Polygraph Program,* DOD Directive, June 1983.

182. U.S. Department of Defense, Department of the Army, *The Validity of Polygraph Testing* (Washington, D.C.: U.S. Government Printing Office, 1983).

183. U.S. Department of Defense, *Present Status of DOD Research on the Polygraph,* report of the DOD Joint Services Group on a Coordinated R&D Program of Lie Detection Research, Aug. 28, 1968.

183a. U.S. Department of Defense, Department of the Air Force, response to OTA survey, 1983.

184. U.S. Department of Justice, background briefer on Presidential Directive, 1983.

185. U.S. Department of Justice, *Presidential Directive on Safeguarding National Security Information,* 1983.

185a. U.S. Department of Justice, "Presidential Directive on Safeguarding National Security Information and Polygraph Examinations of Federal Employees," statement of Richard K. Willard before the House Government Operations Subcommittee on Legislation and National Security, Oct. 19, 1983.

185b. U.S. Department of Justice, response to OTA survey, 1983.

186. U.S. Department of the Treasury, response to OTA survey, 1983.

187. U.S. National Security Agency, initial and supplemental responses to OTA survey, 1983.

188. U.S. Office of Personnel Management, *Federal Personnel Manual System,* ch. 736, app. D, "Use of the Polygraph in Personnel Investigations of Competitive Service Applicants and Appointees to Competitive Service Positions" (Washington, D.C.: U.S. Office of Personnel Management, 1973).

189. *United States* v. *Brown,* 557 F.2d 541, 1977.

190. *United States* v. *Stifel,* 433 F.2d 431 (6th Cir. 1970).

191. *United States* v. *Williams,* 583 F.2d (2nd Cir. 1978).

192. "Uproar Over News Leaks and Lie Detectors," *U.S. News and World Report* 54:8, Aug. 15, 1963.

193. Waid, W. M., "Skin Conductance Response to Both Signaled and Unsignaled Noxious Stimulation Predicts Level of Socialization," *Journal of Personality and Social Psychology* 34, 923-929, 1976.

194. Waid, W. M., and Orne, M. T., "Cognitive, Social, and Personality Processes in the Physiological Detection of Deception," *Advances in Experimental Social Psychology, Berkowitz, L. (ed.)* 14:61-106, 1981.

195. Waid, W. M., and Orne, M. T., "The Physiological Detection of Deception," *American Scientist* 70:402-409, 1982.

196. Waid, W. M., Orne, E. C., Cook, M. R., and Orne, M. T., "Effects of Attention, as Indexed by Subsequent Memory, on Electrodermal Detection of Information," *Journal of Applied Psychology* 63:728-733, 1978.

197. Waid, W. M., Orne, E. C., Cook, M. R., and Orne, M. T., "Meprobamate Reduces Accuracy of Physiological Detection of Deception," *Science* 212:71-73, 1981.

198. Waid, W. M., Orne, M. T., and Wilson, S. K., "Effects of Level of Socialization on Electrodermal Detection of Deception," *Psychophysiology* 16:15-22, 1979.

199. Waid, W. M., Orne, M. T., and Wison, S. K., "Socialization, Awareness and the Electrodermal Response to Deception and Self-Disclosure," *Journal of Abnormal Psychology* 88: 663-666, 1979.

200. Waid, W. M., Wilson, S. K., and Orne, M. T., "Cross-Model Physiological Effects of Electrodermal Lability in the Detection of Deception," *Journal of Personality and Social Psychology* 40:1118-1125, 1981.

201. Weir, R. J., "In Defense of the Relevant-Irrelevant Polygraph Test," *Polygraph* 3:119-166, 1974.
202. Weir, R. J., "Stimulation Procedures—A Conservative View," *Polygraph* 7:209-214, 1978.
203. Weir, R. J., and Atwood, W. F., "Applicant Screening Polygraph Examinations," *Polygraph* 10:129-142, 1981.
204. Whiteside, G., *Profits vs. Polygraph* (Phoenix, Ariz.: Associated Grocers, 1979).
205. Wicklander, D. E., and Hunter, F. L., "The Influence of Auxiliary Sources of Information in Polygraph Diagnoses," *Journal of Police Science and Administration* 3:405-409, 1975.
206. Widacki, J., "Analiza Przestanek Diagnozowania w Badanich Poligraficznych (The Analysis of Diagnostic Premises in Polygraph Examinations)" (Katowice, Poland: Uniwersytetu Slaskiego, 1982).
207. Widacki, J., and Horvath, F., "An Experimental Investigation of the Relative Validity and Utility of the Polygraph Technique and Three Other Common Methods of Criminal Identification," *Journal of Forensic Sciences* 23:596-601, 1978.
208. Ziskin, J., *Coping With Psychiatric and Psychological Testimony* (Venice, Calif.: Law and Psychology Press, 1982).

Office of Technology Assessment

The Office of Technology Assessment (OTA) was created in 1972 as an analytical arm of Congress. OTA's basic function is to help legislative policymakers anticipate and plan for the consequences of technological changes and to examine the many ways, expected and unexpected, in which technology affects people's lives. The assessment of technology calls for exploration of the physical, biological, economic, social, and political impacts that can result from applications of scientific knowledge. OTA provides Congress with independent and timely information about the potential effects—both beneficial and harmful—of technological applications.

Requests for studies are made by chairmen of standing committees of the House of Representatives or Senate; by the Technology Assessment Board, the governing body of OTA; or by the Director of OTA in consultation with the Board.

The Technology Assessment Board is composed of six members of the House, six members of the Senate, and the OTA Director, who is a nonvoting member.

OTA has studies under way in nine program areas: energy and materials; industry, technology, and employment; international security and commerce; biological applications; food and renewable resources; health; communication and information technologies; oceans and environment; and science, transportation, and innovation.

OTA-TM-H-15

JOHN H. GIBBONS
DIRECTOR

Congress of the United States
OFFICE OF TECHNOLOGY ASSESSMENT
WASHINGTON, D.C. 20510

ERRATA ###### ERRATA ###########
FOR: Scientific Validity of Polygraph Testing: A Research Review and Evaluation

"Also serving on the polygraph advisory panel was

Martin T. Orne
Professor of Psychiatry
University of Pennsylvania Medical School"